Resampling Sensitivity of High-Dimensional PCA

Haoyu Wang
Department of Mathematics
Yale University
New Haven, CT 06511
haoyu.wang@yale.edu

Abstract

The stability and sensitivity of statistical methods or algorithms with respect to their data is an essential problem in machine learning and statistics. One fundamental way to measure the stability of an algorithm is to study its performance under resampling of the data. In this paper, we study the resampling sensitivity for the principal component analysis (PCA). When the population covariance matrix of the data does not have a strong spike (i.e. in the subcritical regime), we show that PCA is resampling sensitive by establishing a sharp threshold for the resampling strength, above which resampling even a negligible portion of the input may completely change the principal component; below this threshold, a moderate resampling has almost no effect on the output. In contrast, if the population covariance matrix possesses a strong spike, PCA will be stabilized by the planted signal. All of our results hold with universality, regardless of the underlying data distribution.

1 Introduction

The study of stability and sensitivity of statistical methods and algorithms with respect to the input data is an important task in machine learning and statistics [BE02, EEPK05, MNPR06, HRS16, DHS21]. The notion of stability for algorithms is also closely related to differential privacy [DR14] and generalization error [KN02]. To measure algorithmic stability, one fundamental question is to study the performance of the algorithm under resampling of its input data [BCRT21, KB21]. Originating from the analysis of Boolean functions [BKS99, GS14], resampling sensitivity (also called noise sensitivity) is an important concept in theoretical computer science, which refers to the phenomenon that resampling a small portion of the random input data may lead to decorrelation of the output.

In this work, we study the resampling sensitivity of principal component analysis (PCA). As one of the most commonly used statistical methods, PCA is widely applied for dimension reduction, feature extraction, etc [Joh07, DT11]. It is also used in other fields such as economics [VK06], finance [ASX17], genetics [Rin08], and so on. The impact of noise on PCA is a significant problem in statistics and machine learning, and has been a subject of extensive research. The performance of PCA under the additive or multiplicative independent perturbation of the data matrix has been well studied (see e.g. [BBAP05, BS06, Pau07, BGN11, CLMW11, FWZ18]). However, the influence of resampling on the outcome, as a distinct form of data corruption, remains poorly understood. In this paper, we aim to address this issue for the first time. Here, we emphasize that the resampling of the input data may not have any structure, and the specific resampling procedure is given in the next subsection. Our primary findings reveal that PCA is sensitive to resampling when the population covariance matrix of the data lacks a strong signal. In such cases, even resampling a negligible portion of the data can cause a significant alteration in the resulting principal component, rendering it orthogonal to the original direction. Conversely, when the population covariance of the data possesses a strong spike, the planted signal acts to stabilize PCA.

1.1 Model and Main Results

Let $\mathbf{z}_1, \ldots, \mathbf{z}_n \in \mathbb{R}^p$ be independent random vectors with covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$, i.e. $\mathbb{E}[\mathbf{z}_i\mathbf{z}_i^\top] = \mathbf{\Sigma}$. The sample covariance matrix of the data $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is defined as $\mathbf{H} := \frac{1}{n}\sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i^\top$, and the principal component of the data refers to the unit eigenvector corresponding to the top eigenvalue of the sample covariance matrix. Equivalently, we can rewrite the sample covariance matrix as $\mathbf{H} = (\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top$, where the square root matrix $\Sigma^{1/2}$ is well defined via the spectral decomposition and $\mathbf{X} \in \mathbb{R}^{p\times n}$ is a random matrix whose columns have an isotropic covariance matrix. The assumptions on the data matrix are stated as follows.

Assumption 1. *Let $\mathbf{X} = (\mathbf{X}_{ij})$ be an $p \times n$ data matrix with independent real valued entries with mean 0 and variance n^{-1},*

$$\mathbf{X}_{ij} = n^{-1/2}x_{ij}, \quad \mathbb{E}[x_{ij}] = 0, \quad \mathbb{E}[x_{ij}^2] = 1. \tag{1}$$

Furthermore, we assume the entries x_{ij} have a sub-exponential decay, that is, there exists a constant $\theta > 0$ such that for $u > 1$,

$$\mathbb{P}(|x_{ij}| > u) \leq \theta^{-1}\exp(-u^\theta). \tag{2}$$

Note that we do not require the i.i.d. condition for the data. The sub-exponential decay assumption is mainly for convenience, and other conditions such as the finiteness of a sufficiently high moment would be enough.

Motivated by high-dimensional statistics, we will work in the proportional growth regime $n \asymp p$.

Assumption 2. *Throughout this paper, to avoid trivial eigenvalues, we will be working in the regime*

$$\lim_{n\to\infty} p/n = \xi \in (0, 1) \quad \text{or} \quad p/n \equiv 1.$$

In the case $\lim p/n = 1$, our assumption $p/n \equiv 1$ is due to some technicalities in random matrix theory. Specifically, our proof relies on the delocalization of eigenvectors in the whole spectrum. As one of the major open problems in random matrix theory, delocalization of eigenvectors near the lower spectral edge is not known in the general case with just $\lim p/n = 1$ [AEK14, BEK$^+$14]. The strictly square assumption $p \equiv n$ can be slightly relaxed to $|n - p| = p^{o(1)}$ (see e.g. [Wan22]), but we do not pursue such an extension for simplicity.

For the population covariance matrix Σ, we are interested in the spiked covariance model, which was initiated by Johnstone [Joh01].

$$\mathbf{\Sigma} = \mathbf{I}_p + \sum_{i=1}^r \widetilde{\sigma}_i \mathbf{y}_i\mathbf{y}_i^\top,$$

where r is a fixed integer, the constants $\{\widetilde{\sigma}_i\}_{i=1}^r$ represent strengths of the signals, and $\{\mathbf{y}_i\}_{i=1}^r$ is an orthonormal basis of eigenvectors. It is well known that the BBP phase transition [BBAP05, BGN11, BKYY16] affirms that if $\widetilde{\sigma}_i > \sqrt{\xi}$, then the i-th spike will give rise to an outlier of the spectrum of the sample covariance matrix $\mathbf{H}$. When all $\widetilde{\sigma}_i \leq \sqrt{\xi}$, we call it a weakly spiked model and in particular when $\mathbf{\Sigma} = \mathbf{I}_p$ we call it the null model. On the other hand, if $\widetilde{\sigma}_i > \sqrt{\xi}$ for some i, we call it a strongly spiked model.

In our work, due to technical reasons, we assume that the population covariance matrix is diagonal.

Assumption 3. *The population covariance matrix is diagonal, i.e.* $\mathbf{\Sigma} = \mathsf{diag}(d_1, \cdots, d_p)$ *with constants* $d_1 \geq \cdots \geq d_p \geq 1$*. Moreover, the population covariance matrix differs from the identity by a finite rank, i.e.* $|\{i : d_i \neq 1\}| = r$ *for some fixed integer* r*.*

The reasons for the diagonal assumption on the population covariance are two-fold: (1) our proof hinges on several technical results in random matrix theory such as delocalization of all eigenvectors and Tracy-Widom concentration of the top eigenvalue, etc. Beyond the null model, these results are only known in the case where the population covariance matrix is diagonal [BKYY16, DY18], etc. (2) The diagonal population covariance implies that the entries in each data vector $\mathbf{z}_i$ are independent. In this way, when implementing the resampling procedure, the diagonal assumption makes resampling of the original data $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ equivalent to resampling of the entries in $\mathbf{X}$. This facilitates the model easier to state and the analysis more tractable.

We order the eigenvalues of the sample covariance matrix $\mathbf{H} := (\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top$ as $\lambda_1 \geq \cdots \geq \lambda_p$, and use $\mathbf{v}_i \in \mathbb{R}^p$ to denote the unit eigenvector corresponding to the eigenvalue λ_i. If the context is clear, we just use $\lambda := \lambda_1$ and $\mathbf{v} := \mathbf{v}_1$ to denote the largest eigenvalue and the top eigenvector. We also consider the eigenvalues and eigenvectors of the Gram matrix $\widehat{\mathbf{H}} := (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X})$. Note that $\widehat{\mathbf{H}}$ and $\mathbf{H}$ have the same non-trivial eigenvalues, and the spectrum of $\widehat{\mathbf{H}}$ is given by $\{\lambda_i\}_{i=1}^n$ with $\lambda_{p+1} = \cdots = \lambda_n = 0$. We denote the unit eigenvectors of $\widehat{\mathbf{H}}$ associated with the eigenvalue λ_i by $\mathbf{u}_i \in \mathbb{R}^n$. Writing $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_p] \in \mathbb{R}^{n\times p}$ and $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_p] \in \mathbb{R}^{p\times p}$, then these eigenvectors may be connected by the singular value decomposition of the data matrix $\mathbf{\Sigma}^{1/2}\mathbf{X} = \mathbf{V}\mathbf{S}\mathbf{U}^\top$, where $\mathbf{S} := \mathsf{diag}(\sigma_1, \cdots, \sigma_p)$ with $\sigma_i = \sqrt{\lambda_i}$ corresponds to the singular values. For convenience, we also denote $\sigma := \sigma_1$. And therefore, up to the sign of the eigenvectors, we have

$$(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top\mathbf{v}_i = \sqrt{\lambda_i}\mathbf{u}_i, \quad (\mathbf{\Sigma}^{1/2}\mathbf{X})\mathbf{u}_i = \sqrt{\lambda_i}\mathbf{v}_i.$$

We now describe the resampling procedure. We first emphasize that resampling the data $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ is equivalent to resampling the matrix $\mathbf{X}$ as the diagonal assumption of the population $\mathbf{\Sigma}$ ensures that the entries in a data vector $\mathbf{z}_i$ are independent. For a positive number $k \leq np$, define the random matrix $\mathbf{X}^{[k]}$ in the following way. Let $S_k = \{(i_1, \alpha_1), \cdots, (i_k, \alpha_k)\}$ be a set of k pairs chosen uniformly at random without replacement from the set of all ordered pairs (i, α) of indices with $1 \leq i \leq p$ and $1 \leq \alpha \leq n$. We assume that the set S_k is independent of the entries of $\mathbf{X}$. The entries of $\mathbf{X}^{[k]}$ are given by

$$\mathbf{X}^{[k]}_{i,\alpha} = \begin{cases} \mathbf{X}'_{i,\alpha} & \text{if } (i,\alpha) \in S_k, \\ \mathbf{X}_{i,\alpha} & \text{otherwise,} \end{cases}$$

where $(\mathbf{X}'_{i,\alpha})_{1\leq i\leq p, 1\leq\alpha\leq n}$ are independent random variables, independent of $\mathbf{X}$, and $\mathbf{X}'_{i,\alpha}$ has the same distribution as $\mathbf{X}_{i,\alpha}$. In other words, $\mathbf{X}^{[k]}$ is obtained from $\mathbf{X}$ by resampling k random entries of the matrix, and therefore $\mathbf{X}^{[k]}$ clearly has the same distribution as $\mathbf{X}$. Let $\mathbf{H}^{[k]} := (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})^\top$ be the sample covariance matrix corresponding to the resampled matrix $\mathbf{X}^{[k]}$. Denote the eigenvalues and the corresponding normalized eigenvectors of $\mathbf{H}^{[k]}$ by $\lambda_1^{[k]} \geq \cdots \geq \lambda_p^{[k]}$ and $\mathbf{v}_1^{[k]}, \cdots, \mathbf{v}_p^{[k]}$. When the context is clear, the principal component is just denoted by $\lambda^{[k]}$ and $\mathbf{v}^{[k]}$. Similarly, denote the eigenvector of the Gram matrix $\widehat{\mathbf{H}}^{[k]} := (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[k]})$ associated with the eigenvalue $\lambda_i^{[k]}$ by $\mathbf{u}_i^{[k]}$.

The sensitivity and stability of PCA crucially depend on how strong the planted signal in the spiked covariance is. To measure the strength of the spikes in the population covariance matrix, we define the set of outlier indices

$$\mathcal{O} := \{i : d_i > 1 + \sqrt{\xi}\}.$$

If $\mathcal{O} = \emptyset$, then the model is weakly spiked. On the other hand, if $\mathcal{O} \neq \emptyset$, the eigenvalues with indices in $\mathcal{O}$ will be an outlier.

For the weakly spiked model, with the resampling parameter in two different regimes, we have the following results.

Theorem 1.1 (Weakly spiked model: sensitivity)**.** *Suppose the data profile $\mathbf{X}, \mathbf{\Sigma}$ satisfy Assumptions 1, 2 and 3 with $\mathcal{O} = \emptyset$, and let $\mathbf{X}^{[k]}$ be the resampled matrix defined as above. For any $\epsilon_0 > 0$, if $k \geq n^{5/3+\epsilon_0}$, then the associated principal components are asymptotically orthogonal, i.e.*

$$\lim_{n\to\infty} \mathbb{E}\left|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle\right| = 0, \quad \textit{and} \quad \lim_{n\to\infty} \mathbb{E}\left|\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle\right| = 0. \tag{3}$$

Moreover, in the null model, the threshold for k can be improved to $k \gg n^{5/3}$.

Theorem 1.2 (Weakly spiked model: stability)**.** *Suppose the data profile $\mathbf{X}, \mathbf{\Sigma}$ satisfy Assumptions 1, 2 and 3 with $\mathcal{O} = \emptyset$, and let $\mathbf{X}^{[k]}$ be the resampled matrix defined as above. For any $\epsilon_0 > 0$,*

$$\max_{1\leq k\leq n^{5/3-\epsilon_0}} \min_{s\in\{-1,1\}} \sqrt{n}\|\mathbf{v} - s\mathbf{v}^{[k]}\|_\infty \xrightarrow{\text{prob}} 0, \tag{4}$$

where $\xrightarrow{\text{prob}}$ means convergence in probability. The same result also holds for $\mathbf{u}$ and $\mathbf{u}^{[k]}$.

These two theorems together state that the critical threshold for the resampling strength is of order $k \asymp n^{5/3}$. Note that $n^{5/3}$ compared with the total number of inputs $np \asymp n^2$ is negligible. We show that, above the threshold $n^{5/3}$, resampling even a negligible portion of the data will result in a dramatic change of the resulting principal component, in the sense that the new principal component is asymptotically orthogonal to the old one; while below the threshold, a relatively mild resampling has almost no effect on the corresponding new principal component. If considering the eigenvector overlaps $|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle|$ and $|\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle|$, these quantities exhibit sharp phase transitions from 1 to 0 near the critical threshold $k \asymp n^{5/3}$.

We remark that the phase transition stated in the above theorems is not restricted to the top eigenvectors $\mathbf{v}, \mathbf{v}^{[k]}, \mathbf{u}, \mathbf{u}^{[k]}$. With essentially the same arguments, we can prove that for any fixed $m > 0$, the m-th leading eigenvectors $\mathbf{v}_m, \mathbf{v}_m^{[k]}$ and $\mathbf{u}_m, \mathbf{u}_m^{[k]}$ exhibit the same phase transition at the critical threshold of the same order $n^{5/3}$.

In contrast, for the strongly spiked model, the spike forces the principal components to be correlated with the planted signals. Therefore, in this case, PCA performs better stability than in the weakly spiked model.

Theorem 1.3 (Strongly spiked model: stability)**.** *Suppose the data profile $\mathbf{X}, \mathbf{\Sigma}$ satisfy Assumptions 1, 2 and 3, and let $\mathbf{X}^{[k]}$ be the resampled matrix defined as above. If $\mathcal{O} \neq \emptyset$, for any $i \in \mathcal{O}$ and $k \geq 0$, almost surely we have*

$$\left|\langle \mathbf{v}_i, \mathbf{v}_i^{[k]}\rangle\right| \geq 1 - 4\left(1 - \sqrt{\frac{1 - \frac{\xi}{(d_i-1)^2}}{1 + \frac{\xi}{d_i-1}}}\right) + o(1). \tag{5}$$

Finally, we remark that though the resampling procedure is done uniformly at random for all entries, the proof proceeds as well if we choose to resample the columns. If we resample K columns uniformly at random, then all results are still valid with the threshold $K \asymp n^{2/3}$.

1.2 Comparison with Previous Work

The resampling sensitivity of the leading eigenvector for Wigner matrices and Erdős-Rényi graphs has been studied in [BLZ20, BL22]. The problem discussed in our paper shares a similar prototype, and in particular the threshold $k \asymp n^{5/3}$ for the stability-sensitivity transition in the weakly spiked model coincides with the the threshold for Wigner matrices. Here we want to highlight the differences between our work and theirs.

(i) Our model, the sample covariance matrix, has a Gram structure. This nonlinearity in the matrix model requires more delicate techniques to analyze and the proofs in previous work on symmetric linear model cannot be directly applied here. To overcome the nonlinearity, an important linearization technique is introduced to reduce the interdependency of entries in the sample covariance matrix.

(ii) Due to the Gram matrix structure, the entries in the sample covariance matrix are correlated. In contrast to the case of symmetric matrices, resampling one entry in the data matrix will result in changes of $\Theta(n)$ entries in the sample covariance matrix. Therefore, it is highly non-trivial that our threshold coincides with the threshold for linear models.

(iii) The most important difference: our work is capable of dealing with heteroskedastic data, while previous works are restricted to matrices with identical variances. This makes the applicability of our result much wider. In particular, we establish a clear understanding for the effect of spikes. Such a study of matrix models with planted signals in our work is beyond the scope of previous works.

(iv) From the applied perspective, we uncover connections between the resampling sensitivity phenomenon for PCA with signal detection, differential privacy and database alignment (see Section 4). These application were not addressed in previous works.

Compared with previously mentioned work such as [BBAP05, Pau07, BGN11, FWZ18] that mainly focused on PCA with additive or multiplicative independent noise, our setting is very different. In our model, if writing the resampling effect as an additive or multiplicative perturbation, then this noise is not independent of the signal and does not possess any special structure. In contrast, in previous work, sometimes low-rank assumptions on the structure of the matrix or the noise, or some kind of incoherence conditions were imposed. In our work, we have almost no assumption on the data other than a sub-exponential decay condition. Moreover, we highlight that our results have universality. In particular, we do not need to know the specific distribution of the data and we do not require the data is i.i.d sampled.

A similar framework of PCA with corrupted data is the robust PCA [CLMW11, XCS10]. Regarding connection with robust PCA, our setting does share some similarities with RPCA, as both settings consider corruptions of the original data. In RPCA, corruption is usually related to outlier distribution and we focus on recovering the data. On the other hand, the resampling sensitivity setting in our work studies corruption of data by an independent copy of the same distribution. The key point of our result is that even for two data matrices with the same marginal distribution, a negligible portion of different entries may result in having drastically different principal components.

2 Sensitivity Regime for Weakly Spiked Model

2.1 Heuristics

We provide a heuristic argument for deriving the threshold for the sensitivity regime. We consider the derivative of the top eigenvalue as a function of the matrix entries. For a symmetric matrix $\mathbf{A}$ with an eigenpair $(\lambda, \mathbf{v})$, the derivative of λ with respect to the matrix entries is given by $\mathrm{d}\lambda = \mathbf{v}^\top(\mathrm{d}\mathbf{A})\mathbf{v}$. Motivated by this, we have the approximation

$$\lambda^{[1]} - \lambda \approx \mathbf{v}^\top \left[(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[1]})(\mathbf{\Sigma}^{1/2}\mathbf{X}^{[1]})^\top - (\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top\right]\mathbf{v}.$$

Note that the matrix in the parenthesis has only $\Theta(p)$ non-zero entries, and each entry is roughly of size $O(n^{-1+\varepsilon/2})$ for an arbitrarily small $\varepsilon > 0$ thanks to the sub-exponential decay assumption (2). Also, the eigenvector $\mathbf{v}$ is delocalized in the sense that $|\mathbf{v}(m)| \approx p^{-1/2+\varepsilon/4}$ for all $m = 1, \ldots, p$. A central limit theorem yields that approximately we have

$$\lambda^{[1]} - \lambda \approx O\left(\sqrt{p} n^{-1+\varepsilon/2} p^{-1+\varepsilon/4} p^{-1+\varepsilon/4}\right) = O(n^{-3/2+\varepsilon}).$$

By this heuristic argument and central limit theorem, we have

$$\lambda^{[k]} - \lambda \approx O(\sqrt{k} n^{-3/2+\varepsilon}).$$

Note that from random matrix theory, we know that $\lambda_1 - \lambda_2$ is approximately of order $n^{-2/3}$. Therefore, if we have $\sqrt{k} n^{-3/2+\varepsilon} \ll n^{-2/3}$ (this corresponds to $k \ll n^{5/3-\varepsilon}$), then the difference the two top eigenvalues λ and $\lambda^{[k]}$ is much smaller than the first two eigenvalues λ_1 and λ_2 of the matrix $(\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top$. This implies that the perturbation effect on $\mathbf{X}^{[k]}$ is small, and therefore in this case it is plausible to believe that $\mathbf{v}^{[k]}$ is just a small perturbation of $\mathbf{v}$. Thus, for the threshold of the sensitivity regime, we must expect $k \gg n^{5/3}$.

Our proof is essentially trying to make the above heuristics rigorous. To do this, a key observation is that the inner products $\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle$ and $\langle \mathbf{u}, \mathbf{u}^{[k]} \rangle$ can be related to the variance of the leading eigenvalue.

2.2 Connection with Variance of Top Eigenvalue

As mentioned above, the key step in the proof for sensitivity regime is to establish a connection between the inner products of top eigenvalues with the variance of the top eigenvalue. Specifically, we will prove

$$\mathbb{E}\left[|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|^2\right] \le C \frac{n^3 \mathsf{Var}(\sigma)}{k} + o(1),$$

where $C > 0$ is some universal constant and $\sigma = \sqrt{\lambda}$ is the leading singular value. A similar result is also true for $\mathbf{u}$ and $\mathbf{u}^{[k]}$. More details are deferred to the Supplemental Material.

From random matrix theory [LS16, Wan19], we have $\mathsf{Var}(\sigma) = O(n^{-4/3+\epsilon_0/2})$ for any $\epsilon_0 > 0$. Then, based on this inequality, we derive the threshold $k \ge n^{5/3+\epsilon_0}$ for the sensitivity regime. Moreover, in the null model, it is shown in [LR10] that the variance estimates can be slightly enhanced to $\mathsf{Var}(\sigma) = O(n^{-4/3})$, which results in the improved threshold $k \gg n^{5/3}$ in this case.

3 Stability Regime

3.1 Weakly Spiked Model

To establish the stability of PCA when the resampling strength is mild, we will utilize tools from random matrix theory and specifically the proof relies on the analysis of the resolvent matrix.

Furthermore, to simplify the nonlinearity caused by Gram structure of the sample covariance matrix, when considering the resolvent we use a linearization trick. For any $z \in \mathbb{C}$ with $\operatorname{Im} z > 0$, the resolvent is defined as

$$\mathbf{R}(z) := \begin{pmatrix} -\mathbf{I}_n & (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}) & -z\mathbf{I}_p \end{pmatrix}^{-1}.$$

Similarly, we denote the resolvent of $\mathbf{X}^{[k]}$ by $\mathbf{R}^{[k]}(z)$. The key idea for the proofs in the stability regime is that eigenvectors can be approximated by resolvents and the resolvents are stable under moderate resampling.

Resolvent Approximation To illustrate the usefulness of the resolvent, we show that the entries of the resolvent can be used to approximate the product of entries in the eigenvector. For some small $\delta > 0$, let $z_0 = \lambda + \mathrm{i}\eta$ with $\eta = n^{-2/3-\delta}$. In the regime $k \le n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$, there exists some small $c > 0$ such that for all $i, j = 1, \ldots, p$, we have

$$|\eta \operatorname{Im} \mathbf{R}_{n+i,n+j}(z_0) - \mathbf{v}(i)\mathbf{v}(j)| \le n^{-1-c},$$

and

$$\left|\eta \operatorname{Im} \mathbf{R}^{[k]}_{n+i,n+j}(z_0) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \le n^{-1-c}.$$

A similar result also holds for $\mathbf{u}$ and $\mathbf{u}^{[k]}$, and more details are deferred to the Supplemental Material.

Stability of the Resolvent Since the eigenvector can be approximated by the resolvent, it suffices to show the stability of the resolvent. Consider the regime $k \le n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$. For some small $\delta > 0$ and all $z = E + \mathrm{i}\eta$ that is close to the upper spectral edge and $\eta = n^{-2/3-\delta}$, there exists a small constant $c > 0$ such that the following is true for all $i, j = 1, \ldots, p$ and $\alpha, \beta = 1, \ldots, n$,

$$\left|\mathbf{R}^{[k]}_{\alpha\beta}(z) - \mathbf{R}_{\alpha\beta}(z)\right| \le \frac{1}{n^{1+c}\eta},$$

and

$$\left|\mathbf{R}^{[k]}_{n+i,n+j}(z) - \mathbf{R}_{n+i,n+j}(z)\right| \le \frac{1}{n^{1+c}\eta}.$$

This is the main technical part of the whole argument, and its proof relies on the Lindeberg exchange method and a martingale concentration argument.

Combining the stability of the resolvents with the resolvent approximation for eigenvectors, we can conclude that $\mathbf{v}$ and $\mathbf{v}^{[k]}$ must be close (similarly, also for $\mathbf{u}$ and $\mathbf{u}^{[k]}$).

3.2 Strongly Spiked Model

For strongly spiked model, the famous BBP phase transition [BBAP05, BGN11] states that the top eigenvector $\mathbf{v}$ has non-trivial correlation with the planted signal in the spiked covariance matrix. In our model, we can write $\mathbf{\Sigma} = \sum_{i=1}^p d_i \mathbf{e}_i \mathbf{e}_i^\top$, where $\mathbf{e}_i \in \mathbb{R}^p$ is the i-th coordinate vector which has 1 as the i-th entry and 0 otherwise. With the presence of a strong spike $d_1 > 1 + \sqrt{\xi}$, almost surely, we have

$$\langle \mathbf{v}, \mathbf{e}_1 \rangle^2 = \frac{1 - \frac{\xi}{(d_1-1)^2}}{1 + \frac{\xi}{d_1 - 1}} + o(1).$$

That is, the principal component is forced to lie on a cone around the signal direction $\mathbf{e}_1$. The same phenomenon also happens for the resampled principal component $\mathbf{v}^{[k]}$ for all $k \ge 0$. Since both two principal components lie on the same cone, the angle between them cannot be too large (if choosing an appropriate $\pm$ phase). Hence, we have a non-trivial lower bound for their overlap $|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|$, which confirms that a strong signal helps stabilize PCA.

4 Discussions and Applications

4.1 Further Extensions

A natural direction for future work is to relax the diagonal assumption of the population covariance matrix. For a general population, the entries in each data vector may be correlated. This complicates the resampling procedure. Meanwhile, from the random matrix theory side, some necessary ingredients such as eigenvector delocalization in the whole spectrum and eigenvalue gap property are unknown. A full understanding of the sensitivity of PCA with a general anisotropic population would be an interesting open problem.

4.2 Extensions to Other Statistical Methods

Within the general PCA framework, one important variant is the kernel PCA, which is closely related to the widely used spectral clustering [NJW01, VL07, CBG16]. The corresponding kernel random matrices were studied in [EK10, CS13]. However, the study of these kernel random matrices are far from being well-understood. In particular, the study of eigenvectors were very limited.

It would be interesting to explore whether other statistical method share the same resampling sensitivity phenomenon. Random matrices associated with canonical correlation analysis (CCA) or multivariate analysis of variance (MANOVA) are well studied [HPZ16, HPY18, Yan22]. We anticipate that these models exhibit a similar stability-sensitivity transition as in PCA.

4.3 Signal Detection

Our results have a natural connection with the detection of signals in the spiked covariance model. For a data matrix $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{p\times n}$ in which $\mathbf{z}_i$ are independent random vectors with a rank-one spiked covariance matrix $\mathbf{\Sigma} = \mathsf{diag}(1+\lambda, 1, \ldots, 1)$, our goal is to detect the presence of the spike. To do this, we first subsample $(n - n^{5/6})$ columns uniformly at random to form a submatrix $\mathbf{Z}'$ and denote its principal component as $\mathbf{v}'$. Then we uniformly at random pick $K = n^{5/6}$ samples in the $(n - n^{5/6})$ columns of $\mathbf{Z}'$ and replace them by the remaining $n^{5/6}$ columns in the original data $\mathbf{Z}$. This new submatrix is denoted as $\mathbf{Z}''$ and let $\mathbf{v}''$ be its principal component. Consider the test statistic $\tau := |\langle \mathbf{v}', \mathbf{v}'' \rangle|$. Then by our results Theorem 1.1 and Theorem 1.3, if $\lambda \le \sqrt{\xi}$, we expect $\tau = o(1)$; while if $\lambda > \sqrt{\xi}$, we will have $\tau \ge c$ for some $c > 0$ almost surely.

It is well-known that $\lambda = \sqrt{\xi}$ is the information-theoretic limit for strong detection of spikes [PWBM18, JCL21]. Hence, our resampling sensitivity result yields another test statistic achieving optimal strong detection.

4.4 Differential Privacy

In our paper, we study the stability of the top eigenvalue and the top eigenvector under resampling in terms of bounding the ℓ_∞ distance. Such stability estimates can be regarded as the global sensitivity of PCA performed on neighboring datasets. This measurement is closely related to the analysis of differential privacy [DR14]. PCA under differential privacy was previous studied in [BDMN05, CSS13], etc. Our result revisit the problem of designing a private algorithm for solving the principal component. Here we remark that though the statements in Theorem 1.1 and Theorem

1.2 are qualitative, a careful examination of the proof can yield some quantitative estimates. Based on the stability estimates in terms of the ℓ_∞ metric, a simple Laplace mechanism produces a differentially private version of PCA for computing the top eigenvalue or the top eigenvector. However, compared with [CSS13], our results are limited in the sense that their results are non-asymptotic for all sample size n and data dimension p, while ours are restricted to the proportional growth regime.

Moreover, previous works on differentially private PCA focused on neighbouring datasets that differ by one sample vector. Our result may be seen as a refined notion of privacy, since we can analyze the sensitivity of PCA over two "neighbouring" datasets with k different entries for any k.

Meanwhile, the largest eigenvalue of the sample covariance matrix plays an important rule in hypothesis testing. For example, the Roy's largest root test is used in many problems (see e.g. [JN17]). Our result may provide useful insights to construct a differentially private test statistic based on the top eigenvalue.

4.5 Database Alignment

Database alignment (or in some cases graph matching) refers to the optimization problem in which we are given two datasets and the goal is to find an optimal correspondence between the samples and features that maximally align the data. For datasets $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{p\times n}$, we look for permutations $\pi_{\mathrm{s}} \in \mathcal{S}_n$ and $\pi_{\mathrm{f}} \in \mathcal{S}_p$ to solve the optimization problem

$$\max_{\pi_{\mathrm{s}},\pi_{\mathrm{f}}} \sum_{i=1}^{p}\sum_{\alpha=1}^{n} \mathbf{X}_{i\alpha}\mathbf{Y}_{\pi_{\mathrm{f}}(i)\pi_{\mathrm{s}}(\alpha)},$$

where $\mathcal{S}_n$ and $\mathcal{S}_p$ are the sets of all permutations on $[n]$ and $[p]$, respectively. This problem is closely related to the Quadratic Assignment Problem (QAP), which is known to be NP-hard to solve or even approximate. The study of the alignment problem for correlated random databases has a long history. The previous work mainly considered matrices that are correlated through some additive perturbation, and some of the general model were studied with a homogeneous correlation (i.e. the correlation between all correponding pairs are the same). See for example [DCK19, WXS22] and many other works.

Our resampling procedure may be regarded as a corruption of the dataset, which is a different kind of correlation compared with previous work. To our knowledge, this is the first time to consider database alignment with data corruption. To state the setting of the problem, we have two matrices $\mathbf{X} \in \mathbb{R}^{p\times n}, \quad \mathbf{Y} = \mathbf{\Pi}_{\mathrm{f}}\mathbf{X}^{[k]}\mathbf{\Pi}_{\mathrm{s}}^\top$ where $\mathbf{X}$ is a random matrix satisfying Assumption 1, 2 and 3, and $\mathbf{\Pi}_{\mathrm{s}}$ and $\mathbf{\Pi}_{\mathrm{f}}$ are permutation matrices of order n and p chosen uniformly at random. The goal is to recover the permutations $\mathbf{\Pi}_{\mathrm{s}}$ and $\mathbf{\Pi}_{\mathrm{f}}$ based on the observations $\mathbf{X}$ and $\mathbf{Y}$. Here, we can think of $\mathbf{Y}$ as the unlabeled version of $\mathbf{X}$ with corruption. By considering the covariance matrices, we have

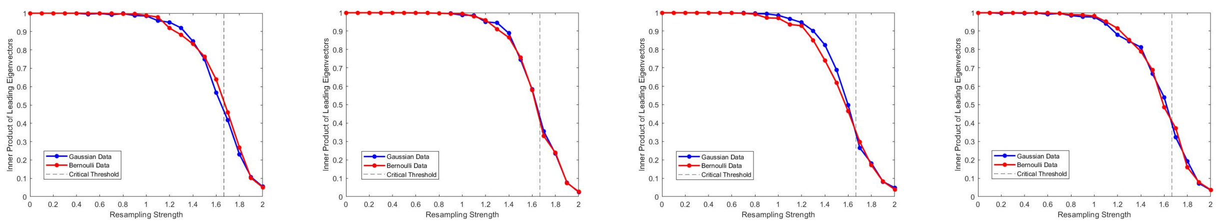
$$\mathbf{A} = \mathbf{X}\mathbf{X}^\top, \quad \mathbf{B} = \mathbf{Y}\mathbf{Y}^\top = \mathbf{\Pi}_{\mathrm{f}}\left(\mathbf{X}^{[k]}(\mathbf{X}^{[k]})^\top\right)\mathbf{\Pi}_{\mathrm{f}}^\top,$$

and similarly for $\mathbf{\Pi}_{\mathrm{s}}$ by considering the Gram matrix. A natural idea to reconstruct the permutations $\mathbf{\Pi}_{\mathrm{s}}$ (and $\mathbf{\Pi}_{\mathrm{f}}$) is to align the top eigenvectors of the matrices $\mathbf{A}$ and $\mathbf{B}$ (and $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$). This spectral method is a natural technique for database alignment and graph matching (e.g. [GLM22]). We are interested in under what resampling strength, this PCA-based algorithm can almost perfectly reconstruct the permutations, and under what condition this method completely fail. See the supplemental material for more discussions.

5 Numerical Experiments

5.1 Synthetic Data

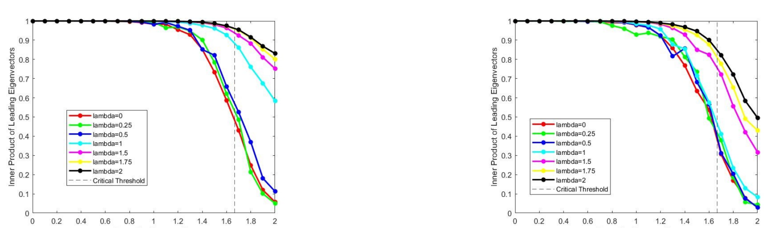
We validate our theoretical prediction by first checking the performance of PCA on synthetic data. We first focus on the weakly spiked model, and consider a data matrix of size 250×1000. To highlight the universality of our results, we will consider Gaussian data and two-point data. In the Gaussian case, the matrix $\mathbf{X}$ consists of i.i.d. $\mathcal{N}(0, 1)$ entries. In the two-point distribution case, the matrix $\mathbf{X}$ consists of i.i.d. entries taking value in $\{\pm 1\}$ with equal probability $\frac{1}{2}$. For the population covariance $\mathbf{\Sigma}$, we consider both the null model and a general weakly spiked model in which the signal is of rank $r = 10$ with strength $\{d_i\}_{i=1}^{10}$ uniformly sampled from $(1, \frac{3}{2})$. To visualize the stability-sensitivity transition, we focus on the overlap of the leading eigenvectors $|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|$ and $|\langle \mathbf{u}, \mathbf{u}^{[k]} \rangle|$ as the observable. Note that, in the stability regime, the asymptotic colinearity (4) implies that $|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|, |\langle \mathbf{u}, \mathbf{u}^{[k]} \rangle| \to 1$. Therefore, we expect a phase transition from 1 to 0 at the critical threshold $k \asymp n^{5/3}$. As shown in Figure 1, there is a clear phase transition for the overlap varying from 1 to 0. It provides good evidence that the transition happens at the critical threshold $k \asymp n^{5/3}$.



(a) The overlap $|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|$ in null model
(b) The overlap $|\langle \mathbf{u}, \mathbf{u}^{[k]} \rangle|$ in null model
(c) The overlap $|\langle \mathbf{v}, \mathbf{v}^{[k]} \rangle|$ in weakly spiked model
(d) The overlap $|\langle \mathbf{u}, \mathbf{u}^{[k]} \rangle|$ in weakly spiked model

Figure 1: Inner products of the leading eigenvectors for 250×1000 matrices with Gaussian and two-point data. The horizontal axis is the resampling strength, given by $\log_n(4k)$. Each experiment is averaged over 50 repetitions.

For the strongly spiked model, we set the population covariance matrix to have a rank-one signal with strength $d_1 = 1 + \lambda$. We consider Gaussian matrices $\mathbf{X}$ as above with sizes 250×1000 and 1000×1000. As shown in Figure 2, the below the corresponding critical signal strength $\lambda = 0.5$ (resp. $\lambda = 1$ for the square case), PCA exhibits stability-sensitivity transition as in the weakly spiked model; while above the critical strength, there is a non-trivial overlap between the principal components. These confirm our theoretical predictions.



(a) Strongly spiked model with $\xi = 1/4$
(b) Strongly spiked model with $\xi = 1$

Figure 2: Overlap of principal components in strongly spiked model with $n = 1000$ samples. The resampling strength is given by $\log_n(k/\xi)$. Each experiment is averaged over 50 repetitions.

5.2 Real Dataset: Peripheral Blood Mononuclear Cells

We further check our results by working with the real dataset of Peripheral Blood Mononuclear Cells (PBMC), where the raw data is publicly available at https://cf.10xgenomics.com/samples/cell/pbmc3k/pbmc3k_filtered_gene_bc_matrices.tar.gz. We use the R toolkit Seurat (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html) for preprocessing of the data (in which the outlier genes will be eliminated). The dataset contains $N = 13714$ gene expressions and $p = 2700$ cells, and PCA is used for dimension reduction and cell clustering. To make it more meaningful in real-world applications, the resampling is done in a slightly different way: we resample a whole gene expression each time instead of resampling single entries of the data matrix. This may be seen as choosing different subsamples when applying a subsampling procedure (which is commonly used in biostatistics). Specifically, we first randomly subsample $n = 3000$ gene expressions and denote it as our data $\mathbf{X}$. Then we resample K gene expressions in $\mathbf{X}$ uniformly at random by replacing them with randomly chosen K gene expressions from the remaining dataset that were not previously subsampled. The data obtained after this resampling is denoted by $\mathbf{Y}$. We then compute the inner product of the corresponding principal component of $\mathbf{X}$ and $\mathbf{Y}$. The results are listed in the following table

$\log(K)/\log(n)$	0.0	0.1	0.2	0.3	0.4	0.5
$\lvert\langle \mathrm{PC}(\mathbf{X}), \mathrm{PC}(\mathbf{Y})\rangle\rvert$	.9999	.9998	.9993	.9989	.9983	.9787
$\log(K)/\log(n)$	0.6	0.7	0.8	0.9	1.0	
$\lvert\langle \mathrm{PC}(\mathbf{X}), \mathrm{PC}(\mathbf{Y})\rangle\rvert$	.9665	.7046	.2772	.1178	.0419	

In this setting, resampling each gene will result in resampling of $p = 2700$ entries in the data. Therefore, the theoretical threshold should be at order $\Theta(n^{5/3}/p) = \Theta(n^{2/3})$, as mentioned at the end of Section 1.1. As we can see from the table, a clear transition happens near this threshold, showing that our theoretical prediction is valid for real data. In this case, the resampling transition can also be interpreted as PCA being sensitive to random subsampling of the data when the data does not contain a strong signal.

A Notations and Organization

We use C to denote generic constant, which may be different in each appearance. We denote $A \lesssim B$ if there exists a universal $C > 0$ such that $A \leq CB$, and denote $A \gtrsim B$ if $A \geq CB$ for some universal $C > 0$. We write $A \asymp B$ if $A \lesssim B$ and $B \lesssim A$.

For the analysis of the sample covariance matrix, it is useful to apply the linearization trick (see e.g. [Tro12, DY18]). Specifically, we will analyze the symmetrization of $\mathbf{\Sigma}^{1/2}\mathbf{X}$. To ease the representation, we drop the dependency on $\mathbf{\Sigma}$ in the notation since the population matrix is fixed throughout, and the symmetrization is denoted as

$$\widetilde{\mathbf{X}} := \begin{pmatrix} 0 & (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}) & 0 \end{pmatrix} \tag{6}$$

The spectrum of the symmetrization $\widetilde{\mathbf{X}}$ are given by the singular values $\{\sqrt{\lambda_m}\}_{m=1}^p$ of $\mathbf{\Sigma}^{1/2}\mathbf{X}$, the symmetrized singular values $\{-\sqrt{\lambda_m}\}_{m=1}^p$, and trivial eigenvalue 0 with multiplicity $n - p$. Let $\mathbf{w}_i := (\mathbf{u}_i^\top, \mathbf{v}_i^\top)^\top \in \mathbb{R}^{n+p}$ be the concatenation of the eigenvectors $\mathbf{u}_i$ and $\mathbf{v}_i$. Then $\mathbf{w}_i$ is the eigenvector of $\widetilde{\mathbf{X}}$ associated with the eigenvalue $\sqrt{\lambda_i}$. Indeed, we have

$$\begin{pmatrix} 0 & (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{pmatrix} = \begin{pmatrix} (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \mathbf{v}_i \\ (\mathbf{\Sigma}^{1/2}\mathbf{X})\mathbf{u}_i \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_i}\mathbf{u}_i \\ \sqrt{\lambda_i}\mathbf{v}_i \end{pmatrix}.$$

An important probabilistic concept that will be used repeatedly is the notion of overwhelming probability.

Definition A.1 (Overwhelming probability)**.** Let $\{\mathcal{E}_N\}$ be a sequence of events. We say that $\mathcal{E}_N$ holds with overwhelming probability if for any (large) $D > 0$, there exists $N_0(D)$ such that for all $N \geq N_0(D)$ we have

$$\mathbb{P}(\mathcal{E}_N) \geq 1 - N^{-D}.$$

Organization The supplemental material is organized as follows. In Section B, we collect some useful tools for the proof, including a variance formula for resampling and classical results from random matrix theory. In Section C, we prove the sensitivity of PCA under excessive resampling for the weakly spiked model. In Section D, we prove that PCA is stable in the weakly spiked model if resampling of the data is moderate. The proof for the strongly spiked model is provided in Section E. Finally, in Section F, we discuss the database alignment problem raised in the Main Part in more detail.

B Preliminaries

B.1 Variance formula and resampling

An essential technique for our proof is the formula for the variance of a function of independent random variables. This formula represents the variance via resampling of the random variables. This idea is first due to Chatterjee [Cha05], and in this paper we will use a slight extension of it as in [BLZ20].

Let $X_1, \cdots, X_N$ be independent random variables taking values in some set $\mathcal{X}$, and let $f : \mathcal{X}^N \to \mathbb{R}$ be some measurable function. Let $X = (X_1, \cdots, X_N)$ and X' be an independent copy of X. We denote

$$X^{(i)} = (X_1, \cdots, X_{i-1}, X_i', X_{i+1}, \cdots, X_N), \quad \text{and} \quad X^{[i]} = (X_1', \cdots, X_i', X_{i+1}, \cdots, X_N).$$

And in general, for $A \subset [N]$, we define X^A to be the random vector obtained from X by replacing the components indexed by A by corresponding components of X'. By a result of Chatterjee [Cha05], we have the following variance formula

$$\mathsf{Var}\left(f(X)\right) = \frac{1}{2}\sum_{i=1}^{N} \mathbb{E}\left[\left(f(X) - f(X^{(i)})\right)\left(f(X^{[i-1]}) - f(X^{[i]})\right)\right].$$

We remark that this variance formula does not depend on the order of the random variables. Therefore, we can consider an arbitrary permutation of $[N]$. Specifically, let $\pi = (\pi(1), \cdots, \pi(N))$ be a random permutation sampled uniformly from the symmetric group $\mathcal{S}_N$ and denote $\pi([i]) := \{\pi(1), \cdots, \pi(i)\}$. Then we have

$$\mathsf{Var}\left(f(X)\right) = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[\left(f(X) - f(X^{(\pi(i))})\right)\left(f(X^{\pi([i-1])}) - f(X^{\pi([i])})\right)\right].$$

Note that, in the formula above, the expectation is taken with respect to both X, X' and the random permutation π.

Let j have uniform distribution over $[N]$. Let $X^{(j)\circ\pi([i-1])}$ denote the vector obtained from $X^{\pi([i-1])}$ by replacing its j-th component by another independent copy of the random variable X_j in the following way: If j belongs to $\pi([i-1])$, then we replace X'_j by X''_j; if j is not in $\pi([i-1])$, then we replace X_j by X'''_j, where X'' and X''' are independent copies of X such that (X, X', X'', X''') are independent. With this notation, we have the following estimates.

Lemma B.1 (Lemma 3 in [BLZ20])**.** *Assume that j is chosen uniformly at random from the set $[N]$ and independently of other random variables involved, we have for any $k \in [N]$,*

$$B_k \leq \frac{2\mathsf{Var}\,(f(X))}{k}\left(\frac{N+1}{N}\right)$$

where for any $i \in [N]$,

$$B_i := \mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\pi([i-1])}) - f(X^{(j)\circ\pi([i-1])})\right)\right]$$

and the expectation is taken with respect to components of vectors, random permutations π and the random variable j.

B.2 Tools from random matrix theory

In this section we collect some classical results in random matrix theory, which will be indispensable for proving the main theorems. These include concentration of the top eigenvalue, eigenvalue rigidity estimates, and eigenvector delocalization. We focus on the *weakly spiked model*, and the BBP phase transition for the strongly spiked model will be deferred to Section E.

To begin with, we first state some basic settings and notations. It is well known that the empirical distribution of the spectrum of the null model (i.e. $\mathbf{\Sigma} = \mathbf{I}$) converges to the Marchenko-Pastur distribution

$$\rho_{\mathsf{MP}}(x) = \frac{1}{2\pi\xi}\sqrt{\frac{[(x-\lambda_-)(\lambda_+ - x)]_+}{x^2}}, \tag{7}$$

where the endpoints of the spectrum are given by

$$\lambda_\pm = (1 \pm \sqrt{\xi})^2. \tag{8}$$

Beyond the null model, where the population covariance matrix is not identity, the convergence of the empirical spectral measure deviates from the ordinary Marchenko-Pastur law in general and instead converges to a distinct limiting distribution known as the deformed Marchenko-Pastur law. The deformed Marchenko-Pastur distribution, denoted as ρ_{fc}, is characterized by its Stieltjes transform. For a probability measure ρ on the real line, the Stieltjes transofrm is defined as

$$m_\rho(z) := \int_{\mathbb{R}} \frac{1}{x-z}\mathrm{d}\rho(x), \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ is the upper half plane of $\mathbb{C}$. The Stieltjes transform is an important object in probability theory with two useful applications: (1) the convergence of the probability measure is equivalent to the convergence of the Stieltjes transform; (2) if the probability measure ρ is absolutely continuous with respect to the Lebesgue measure, it can be recovered from its Stieltjes transform by the inversion formula

$$\rho(x) = \lim_{\eta\downarrow 0}\frac{1}{\pi}\operatorname{Im} m_\rho(x + \mathrm{i}\eta).$$

For the deformed Marchenko-Pastur law, we denote its Stieltjes transform by m_{fc}. The notation m_{fc} stands for free convolution, as the spectral measure of the sample covariance matrix is given by the multiplicative free convolution of the Marchenko-Pastur law and the spectral measure of the population covariance matrix. Suppose the empirical spectral measure $\widehat{\mu}$ of the population covariance matrix $\mathbf{\Sigma}$ converges to a limiting law μ. Then m_{fc} is determined by the following self-consistent equation

$$\frac{1}{m_{\mathsf{fc}}(z)} = -z + \xi \int \frac{t}{1 + m_{\mathsf{fc}}(z)t} d\mu(t). \tag{9}$$

Recall that $\mathbf{\Sigma} = \mathsf{diag}(d_1, \cdots, d_p)$, and indices with $d_i > 1 + \sqrt{\xi}$ are called outliers $i \in \mathcal{O}$. For the weakly spiked model, we have $\mathcal{O} = \emptyset$ and in this case the support of the deformed Marchenko-Pastur law has only one connected component (see e.g. [LS16, DY18]). The right endpoint of the spectrum λ_{R}, which is closely related to the concentration of the top eigenvalue, is determined by

$$\lambda_{\mathrm{R}} = \frac{1}{a}\left(1 + \xi \int \frac{ta}{1 - ta} \mathrm{d}\mu(t)\right), \tag{10}$$

where $a \geq 0$ is the unique solution of the equation

$$\int \left(\frac{ta}{1 - ta}\right)^2 \mathrm{d}\mu(t) = \xi^{-1}.$$

The left endpoint of the spectrum, denoted as λ_{L}, can be determined via a similar way.

An important result in random matrix theory is that the eigenvalues are concentrated. To state the result, we define the typical locations of the eigenvalues:

$$\gamma_m := \inf\left\{E > 0 : \int_{-\infty}^{E} \rho_{\mathsf{fc}}(x)\mathrm{d}x \geq \frac{m}{p}\right\}, \quad 1 \leq m \leq p.$$

A classical result in random matrix theory is the rigidity estimates of the eigenvalues [PY14, BEK+14, LS16, DY18]. Let $\widehat{m} := \min(m, p + 1 - k)$, for any small $\varepsilon > 0$ and large $D > 0$ there exists $n_0(\varepsilon, D)$ such that the following holds for any $n \geq n_0$,

$$\mathbb{P}\left(|\lambda_m - \gamma_m| \leq n^{-\frac{2}{3}+\varepsilon}(\widehat{m})^{-\frac{1}{3}} \text{ for all } 1 \leq m \leq p\right) > 1 - n^{-D}. \tag{11}$$

We remark that the square case $\xi \equiv 1$ is actually significantly different, due to the singularity of the Marchenko-Pastur law at $x = 0$. Near the left spectral edge, the typical eigenvalue spacing would be of order $O(n^{-2})$, which leads to a stronger concentration. In this case, the tight rigidity was proved in [AEK14], and see [Wan22] for more explanations. However, the estimate (11) is good enough for our purpose.

Another important result is the Tracy-Widom limit for the top eigenvalue (see e.g. [PY14, DY18, Wan19, SX21]). Specifically,

Lemma B.2. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. For any $\varepsilon > 0$, with overwhelming probability, we have*

$$|\lambda - \lambda_{\mathrm{R}}| \leq n^{-2/3+\varepsilon}, \quad \textit{and} \quad \mathsf{Var}(\lambda) \leq n^{-4/3+\varepsilon}.$$

In particular, for the null model, we have $\mathsf{Var}(\lambda) \lesssim n^{-4/3}$. Moreover, for any $\delta > 0$, there exists a constant $c_0 > 0$ such that

$$\mathbb{P}\left(\lambda_1 - \lambda_2 \geq c_0 n^{-2/3}\right) \geq 1 - \delta.$$

Proof. The first result follows from the well-known Tracy-Widom limit for the top eigenvalue. Specifically,

$$\lim_{n\to\infty} \mathbb{P}\left(\gamma n^{2/3}(\lambda - \lambda_{\mathrm{R}}) \le s\right) = F_1(s),$$

where γ is a constant depending only on the ratio ξ and the limiting spectral measure μ of the population covariance matrix, and F_1 is the type-1 Tracy-Widom distribution (in particular, [Wan19, SX21] provided quantitative rate of convergence). The variance estimate is then a natural consequence. For the improved variance bound of the null model, the Gaussian case (i.e. the white Wishart ensemble) was proved in [LR10], and the general case follows from universality, i.e. for any fixed m

$$\lim_{n\to\infty} \mathbb{P}\left(\left(n^{2/3}(\lambda_\ell - \lambda_{\mathrm{R}}) \le s_\ell\right)_{1\le\ell\le m}\right) = \lim_{n\to\infty} \mathbb{P}^G\left(\left(n^{2/3}(\lambda_\ell - \lambda_{\mathrm{R}}) \le s_\ell\right)_{1\le\ell\le m}\right),$$

where $\mathbb{P}^G$ denotes the probability measure associated with the Gaussian matrix. The spectral gap estimate also follows from universality that the spectral statistics of the sample covariance matrix is the same as the Gaussian Orthogonal Ensemble (GOE), i.e. for any fixed m

$$\lim_{n\to\infty} \mathbb{P}\left(\left(\gamma n^{2/3}(\lambda_\ell - \lambda_{\mathrm{R}}) \le s_\ell\right)_{1\le\ell\le m}\right) = \lim_{n\to\infty} \mathbb{P}\left(\left(n^{2/3}(\lambda_\ell^{GOE} - 2) \le s_\ell\right)_{1\le\ell\le m}\right)$$

For GOE, the desired spectral gap estimate was proved in e.g. [AGZ10]. □

Moreover, an estimate on the eigenvalue gap near the spectral edge is needed. The following result was proved in [TV12, Wan12]

Lemma B.3. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. For any $c > 0$, there exists $\kappa > 0$ such that for every $1 \le i \le p$, with probability at least $1 - n^{-\kappa}$, we have*

$$|\lambda_i - \lambda_{i+1}| \ge n^{-1-c}.$$

The property of eigenvectors is also a key ingredient for our proof. In particular, we extensively rely on the following delocalization property, which implies that the eigenvectors are distributed roughly uniformly on the unit sphere (see e.g. [PY14, BEK+14, DY18]). This is one of the most significant difference between the weakly spiked model and the strongly spiked model. With a strong spike in the population covariance matrix, the top eigenvector will be forced to lie on a cone around the signal and hence is localized in some sense.

Lemma B.4. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. For any $\varepsilon > 0$, with overwhelming probability, we have*

$$\max_{1\le i\le p} \|\mathbf{v}_i\|_\infty + \max_{1\le j\le p} \|\mathbf{u}_j\|_\infty \le n^{-\frac{1}{2}+\varepsilon}.$$

C Proofs for the Sensitivity Regime of Weakly Spiked Model

C.1 Sensitivity analysis for neighboring data matrices

As a first step, we will first show that resampling of a single entry has little perturbation effect on the top eigenvectors in the weakly spiked model. This will be helpful to control the single entry resampling term in the variance formula (see Lemma B.1).

For any fixed $1 \le i \le p$ and $1 \le \alpha \le n$, let $\mathbf{X}_{(i,\alpha)}$ be the matrix obtained from $\mathbf{X}$ by replacing the (i,α) entry $\mathbf{X}_{i\alpha}$ with $\mathbf{X}'_{i\alpha}$. Define the corresponding covariance matrix $\mathbf{H}_{(i,\alpha)} := (\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})(\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})^\top$, and use $\mathbf{v}^{(i,\alpha)}$ to denote its unit top eigenvector. Similarly, we denote by $\mathbf{u}^{(i,\alpha)}$ the unit top eigenvector of $\widehat{\mathbf{H}}_{(i,\alpha)} := (\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})$.

Lemma C.1. *Let $c > 0$ small and $0 < \delta < \frac{1}{2} - c$. For all $1 \leq i \leq n$ and $1 \leq \alpha \leq p$, on the event $\{\lambda_1 - \lambda_2 \geq n^{-1-c}\}$, with overwhelming probability*

$$\max_{i,\alpha} \min_{s \in \{\pm 1\}} \|\mathbf{v} - s\mathbf{v}^{(i,\alpha)}\|_\infty \leq n^{-\frac{1}{2}-\delta} \tag{12}$$

and similarly

$$\max_{i,\alpha} \min_{s \in \{\pm 1\}} \|\mathbf{u} - s\mathbf{u}^{(i,\alpha)}\|_\infty \leq n^{-\frac{1}{2}-\delta}$$

Proof. Let $\lambda_1^{(i,\alpha)} \geq \cdots \geq \lambda_p^{(1,\alpha)}$ denote the eigenvalues of the matrix $\mathbf{H}_{(i,\alpha)}$, and let $\mathbf{v}_j^{(i,\alpha)}$ denote the unit eigenvector associated with the eigenvalue $\lambda_j^{(i,\alpha)}$. Similarly, we define the unit eigenvectors $\{\mathbf{u}_\beta^{(i,\alpha)}\}$ for the matrix $\widehat{\mathbf{H}}_{(i,\alpha)}$. Using the variational characterization of the eigenvalues, we have

$$\lambda_1 \geq \langle \mathbf{v}_1^{(i,\alpha)}, \mathbf{H}\mathbf{v}_1^{(i,\alpha)} \rangle = \lambda_1^{(i,\alpha)} + \langle \mathbf{v}_1^{(i,\alpha)}, (\mathbf{H} - \mathbf{H}_{(i,\alpha)})\mathbf{v}_1^{(i,\alpha)} \rangle.$$

Recall that $\mathbf{\Sigma} = \mathsf{diag}(d_1, \ldots, d_p)$. Since $\mathbf{X}$ and $\mathbf{X}_{(i,\alpha)}$ differ only at the (i, α) entry, we have

$$\begin{aligned}(\mathbf{H} - \mathbf{H}_{(i,\alpha)})_{j\ell} &= ((\mathbf{\Sigma}^{1/2}\mathbf{X})(\mathbf{\Sigma}^{1/2}\mathbf{X})^\top - (\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})(\mathbf{\Sigma}^{1/2}\mathbf{X}_{(i,\alpha)})^\top)_{j\ell} \\ &= \begin{cases} d_i(\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})\mathbf{X}_{\ell\alpha} & \text{if } j = i, \ell \neq i, \\ d_i(\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})\mathbf{X}_{j\alpha} & \text{if } j \neq i, \ell = i, \\ d_i(\mathbf{X}_{i\alpha}^2 - (\mathbf{X}'_{i\alpha})^2) & \text{if } j = i, \ell = i, \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

Thus, setting $\Delta_{i\alpha} := (\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})$, we have

$$\begin{aligned}&\langle \mathbf{v}_1^{(i,\alpha)}, (\mathbf{H} - \mathbf{H}_{(i,\alpha)})\mathbf{v}_1^{(i,\alpha)} \rangle \\ &= 2\mathbf{v}_1^{(i,\alpha)}(i) d_i \Delta_{i\alpha} \left(\sum_{j=1}^p (\mathbf{X}_{(i,\alpha)})_{j\alpha} \mathbf{v}_1^{(i,\alpha)}(j) - \mathbf{X}'_{i\alpha}\mathbf{v}_1^{(i,\alpha)}(i) \right) + d_i \left(\mathbf{v}_1^{(i,\alpha)}(i) \right)^2 (\mathbf{X}_{i\alpha}^2 - (\mathbf{X}'_{i\alpha})^2) \\ &= 2\mathbf{v}_1^{(i,\alpha)}(i) d_i \Delta_{i\alpha} \left(\mathbf{X}_{(i,\alpha)}^\top \mathbf{v}_1^{(i,\alpha)} \right)(\alpha) + d_i \left(\mathbf{v}_1^{(i,\alpha)}(i) \right)^2 \left(\mathbf{X}_{i\alpha}^2 - (\mathbf{X}'_{i\alpha})^2 - 2(\mathbf{X}_{i\alpha} - \mathbf{X}'_{i\alpha})\mathbf{X}'_{i\alpha} \right) \\ &= 2d_i \sqrt{\lambda_1^{(i,\alpha)}} \mathbf{v}_1^{(i,\alpha)}(i) \mathbf{u}_1^{(i,\alpha)}(\alpha) \Delta_{i\alpha} + d_i \left(\mathbf{v}_1^{(i,\alpha)}(i)^2 \right) \Delta_{i\alpha}^2.\end{aligned} \tag{13}$$

This gives us

$$\lambda_1 \geq \lambda_1^{(i,\alpha)} - 2d_i \sqrt{\lambda_1^{(i,\alpha)}} \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right) \|\mathbf{v}_1^{(i,\alpha)}\|_\infty \|\mathbf{u}_1^{(i,\alpha)}\|_\infty - d_i \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right)^2 \|\mathbf{v}_1^{(i,\alpha)}\|_\infty^2. \tag{14}$$

Similarly,

$$\lambda_1^{(i,\alpha)} \geq \lambda_1 - 2d_i \sqrt{\lambda_1} \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right) \|\mathbf{v}_1\|_\infty \|\mathbf{u}_1\|_\infty - d_i \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}| \right)^2 \|\mathbf{v}_1\|_\infty^2. \tag{15}$$

By Assumption 1, the sub-exponential decay implies $|\mathbf{X}_{i\alpha}|, |\mathbf{X}'_{i\alpha}| \leq n^{-1/2+\varepsilon}$ with overwhelming probability for any $\varepsilon > 0$. Also, by Assumption 3, we know that $d_i = \Theta(1)$. Moreover, by the delocalization of eigenvectors, with overwhelming probability, we have

$$\max \left(\|\mathbf{v}_1\|_\infty, \|\mathbf{u}_1\|_\infty, \|\mathbf{v}_1^{(i,\alpha)}\|_\infty, \|\mathbf{u}_1^{(i,\alpha)}\|_\infty \right) \leq n^{-\frac{1}{2}+\varepsilon}.$$

Moreover, by the rigidity estimates (11), with overwhelming probability we have

$$|\lambda_1 - \lambda_{\mathrm{R}}| \le n^{-\frac{2}{3}+\varepsilon}, \quad |\lambda_1^{(i,\alpha)} - \lambda_{\mathrm{R}}| \le n^{-\frac{2}{3}+\varepsilon}$$

Therefore, combining with a union bound, the above two inequalities (14) and (15) together yield

$$\max_{1\le i\le n, 1\le \alpha\le p} |\lambda_1 - \lambda_1^{(i,\alpha)}| \le n^{-3/2+3\varepsilon} \tag{16}$$

with overwhelming probability.

We write $\mathbf{v}_1^{(i,\alpha)}$ in the orthonormal basis of eigenvectors $\{\mathbf{v}_j\}$:

$$\mathbf{v}_1^{(i,\alpha)} = \sum_{j=1}^{p} a_j \mathbf{v}_j.$$

Using this representation and the spectral theorem,

$$\sum_{j=1}^{p} \lambda_j a_j \mathbf{v}_j = \mathbf{H}\mathbf{v}_1^{(i,\alpha)} = \left(\mathbf{H} - \mathbf{H}_{(i,\alpha)}\right)\mathbf{v}_1^{(i,\alpha)} + \left(\lambda_1^{(i,\alpha)} - \lambda_1\right)\mathbf{v}_1^{(i,\alpha)} + \lambda_1 \mathbf{v}_1^{(i,\alpha)}.$$

As a consequence,

$$\lambda_1 \mathbf{v}_1^{(i,\alpha)} = \sum_{j=1}^{p} \lambda_j a_j \mathbf{v}_j + \left(\mathbf{H}_{(i,\alpha)} - \mathbf{H}\right)\mathbf{v}_1^{(i,\alpha)} + \left(\lambda_1 - \lambda_1^{(i,\alpha)}\right)\mathbf{v}_1^{(i,\alpha)}.$$

For $j \ne 1$, taking inner product with $\mathbf{v}_j$ yields

$$\lambda_1 a_j = \langle \mathbf{v}_j, \lambda_1 \mathbf{v}_1^{(i,\alpha)}\rangle = \lambda_j a_j + \langle \mathbf{v}_j, (\mathbf{H}_{(i,\alpha)} - \mathbf{H})\mathbf{v}_1^{(i,\alpha)}\rangle + (\lambda_1 - \lambda_1^{(i,\alpha)})a_j,$$

which implies

$$\left((\lambda_1 - \lambda_j) + (\lambda_1^{(i,\alpha)} - \lambda_1)\right) a_j = \langle \mathbf{v}_j, (\mathbf{H}_{(i,\alpha)} - \mathbf{H})\mathbf{v}_1^{(i,\alpha)}\rangle. \tag{17}$$

By a similar computation as in (13), we have

$$\begin{aligned} \left|\langle \mathbf{v}_j, (\mathbf{H}_{(i,\alpha)} - \mathbf{H})\mathbf{v}_1^{(i,\alpha)}\rangle\right| &= \left| d_i \Delta_{i\alpha} \left(\sqrt{\lambda_1^{(i,\alpha)}}\mathbf{v}_j(i)\mathbf{u}_1^{(i,\alpha)}(\alpha) + \sqrt{\lambda_\beta}\mathbf{v}_1^{(i,\alpha)}(i)\mathbf{u}_j(\alpha)\right)\right| \\ &\lesssim \left(|\mathbf{X}_{i\alpha}| + |\mathbf{X}'_{i\alpha}|\right)\left(\|\mathbf{v}_j\|_\infty \|\mathbf{u}_1^{(i,\alpha)}\|_\infty + \|\mathbf{v}_1^{(i,\alpha)}\|_\infty \|\mathbf{u}_j\|_\infty\right) \\ &\le n^{-\frac{3}{2}+3\varepsilon} \end{aligned} \tag{18}$$

with overwhelming probability, where the second step follows from rigidity of eigenvalues and the last step follows from the sub-exponential decay assumption and delocalization of eigenvectors.

Consider the event $\mathcal{E} := \{\lambda_1 - \lambda_2 \ge n^{-1-c}\}$. Fix some $\omega > 0$ small. By rigidity of eigenvalues (11), on the event $\mathcal{E}$, with overwhelming probability

$$\lambda_1 - \lambda_j \gtrsim \begin{cases} n^{-1-c} & \text{if } \ 2 \le j \le n^\omega, \\ j^{2/3} n^{-2/3} & \text{if } \ n^\omega < j \le p. \end{cases} \tag{19}$$

On the event $\mathcal{E}$, using (16), (18) and (19), with overwhelming probability we have

$$|a_j| \lesssim \begin{cases} n^{-\frac{1}{2}+c+3\varepsilon} & \text{if } \ 2 \le j \le n^\omega, \\ j^{-\frac{2}{3}} n^{-\frac{5}{6}+3\varepsilon} & \text{if } \ n^\omega < j \le p. \end{cases} \tag{20}$$

Choose $s = a_1/|a_1|$. Note that $1 - |a_1| \le \sum_{j=2}^{p} |a_j|$. Thanks to the delocalization of eigenvectors, with overwhelming probability, we have

$$\|s\mathbf{v}_1 - \mathbf{v}_1^{(i,\alpha)}\|_\infty = \left\|(s - a_1)\mathbf{v}_1 + \sum_{j=2}^{p} a_j \mathbf{v}_j\right\|_\infty \le (1 - |a_1|)\|\mathbf{v}_1\|_\infty + \sum_{j=2}^{p} |a_j| \|\mathbf{v}_j\|_\infty \le n^{-\frac{1}{2}+\varepsilon} \sum_{j=2}^{p} |a_j|.$$

Thus, on the event $\mathcal{E}$, it follows from (19) that

$$\begin{aligned}\|s\mathbf{v}_1 - \mathbf{v}_1^{(i,\alpha)}\|_\infty &\lesssim n^{-\frac{1}{2}+\varepsilon}\left(n^{-\frac{1}{2}+3\varepsilon+c+\omega} + n^{-\frac{5}{6}+3\varepsilon}\sum_{j=n^\omega}^{p} j^{-\frac{2}{3}}\right)\\ &\lesssim n^{-1+4\varepsilon+c+\omega} + n^{-1+4\varepsilon}.\end{aligned}$$

Choosing ε and ω small enough so that $4\varepsilon + c + \omega < \frac{1}{2}$, we conclude that (12) is true.

A similar bound for $\mathbf{u}$ can be shown by the same arguments for $\widehat{\mathbf{H}} = (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top(\mathbf{\Sigma}^{1/2}\mathbf{X})$. Hence, we have shown the desired results. □

C.2 Proof of Theorem 1.1

Now we are ready to prove the resampling sensitivity for the weakly spiked model.

Let $\mathbf{X}'' \in \mathbb{R}^{p\times n}$ be a copy of $\mathbf{X}$ that is independent of $\mathbf{X}$ and $\mathbf{X}'$. For an arbitrary index (i,α) with $1 \le i \le p$ and $1 \le \alpha \le n$, we introduce another random matrix $\mathbf{Y}_{(i,\alpha)}$ obtained from $\mathbf{X}$ by replacing the (i,α) entry $\mathbf{X}_{i\alpha}$ by $\mathbf{X}''_{i\alpha}$. Similarly, we denote $\mathbf{Y}^{[k]}_{(i,\alpha)}$ the matrix obtained via the same procedure from $\mathbf{X}^{[k]}$. For the matrix $\mathbf{X}^{[k]}$, we do the similar resampling procedure in the following way: if $(i,\alpha) \in S_k$, then replace $\mathbf{X}^{[k]}_{i\alpha}$ with $\mathbf{X}''_{i\alpha}$; if $(i,\alpha) \notin S_k$, then replace $\mathbf{X}^{[k]}_{i\alpha}$ with $\mathbf{X}'''_{i\alpha}$, where $\mathbf{X}'''$ is another independent copy of $\mathbf{X}$, $\mathbf{X}'$ and $\mathbf{X}''$.

In the following analysis, we choose an index (s,θ) uniformly at random from the set of all pairs $\{(i,\alpha) : 1 \le i \le p, 1 \le \alpha \le n\}$. Let μ be the top singular value of $\mathbf{Y}_{(s,\theta)}$ and use $\mathbf{f} \in \mathbb{R}^p$ and $\mathbf{g} \in \mathbb{R}^n$ to denote the normalized top left and right singular vectors of $\mathbf{Y}_{(s,\theta)}$. Similarly, we define $\mu^{[k]}$, $\mathbf{f}^{[k]}$ and $\mathbf{g}^{[k]}$ for $\mathbf{Y}^{[k]}_{(s,\theta)}$. We also denote by $\mathbf{h}$ and $\mathbf{h}^{[k]}$ the concatenation of $\mathbf{f}, \mathbf{g}$ and $\mathbf{f}^{[k]}, \mathbf{g}^{[k]}$, respectively. Finally, let $\widetilde{\mathbf{X}}^{[k]}$, $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}^{[k]}$ be the symmetrization (6) of $\mathbf{X}^{[k]}$, $\mathbf{Y}$ and $\mathbf{Y}^{[k]}$, respectively. When the context is clear, we will omit the index (s,θ) for the convenience of notations.

Step 1. By Lemma B.1, we have

$$\frac{2\mathsf{Var}(\sigma)}{k} \cdot \frac{np+1}{np} \ge \mathbb{E}\left[(\sigma - \mu)\left(\sigma^{[k]} - \mu^{[k]}\right)\right]. \tag{21}$$

Using the variational characterization of the top singular value

$$\langle \mathbf{f}, \mathbf{\Sigma}^{1/2}\mathbf{X}\mathbf{g}\rangle \le \sigma = \langle \mathbf{v}, \mathbf{\Sigma}^{1/2}\mathbf{X}\mathbf{u}\rangle, \quad \langle \mathbf{v}, \mathbf{\Sigma}^{1/2}\mathbf{Y}\mathbf{u}\rangle \le \mu = \langle \mathbf{f}, \mathbf{\Sigma}^{1/2}\mathbf{Y}\mathbf{g}\rangle.$$

This implies

$$\langle \mathbf{f}, \mathbf{\Sigma}^{1/2}(\mathbf{X}-\mathbf{Y})\mathbf{g}\rangle \le \sigma - \mu \le \langle \mathbf{v}, \mathbf{\Sigma}^{1/2}(\mathbf{X}-\mathbf{Y})\mathbf{u}\rangle. \tag{22}$$

Applying the same arguments to $\mathbf{X}^{[k]}$ and $\mathbf{Y}^{[k]}$, we have

$$\left\langle \mathbf{f}^{[k]}, \mathbf{\Sigma}^{1/2}\left(\mathbf{X}^{[k]} - \mathbf{Y}^{[k]}\right)\mathbf{g}^{[k]}\right\rangle \le \sigma^{[k]} - \mu^{[k]} \le \left\langle \mathbf{v}^{[k]}, \mathbf{\Sigma}^{1/2}\left(\mathbf{X}^{[k]} - \mathbf{Y}^{[k]}\right)\mathbf{u}^{[k]}\right\rangle.$$

Since the matrices $\mathbf{X}$ and $\mathbf{Y}$ differ only at the (s,θ) entry, for any vectors $\mathbf{a}\in\mathbb{R}^p$ and $\mathbf{b}\in\mathbb{R}^n$ we have

$$\langle \mathbf{a}, \boldsymbol{\Sigma}^{1/2}(\mathbf{X}-\mathbf{Y})\mathbf{b}\rangle = \sqrt{d_s}\Delta_{s\theta}\,\mathbf{a}(s)\mathbf{b}(\theta), \quad \left\langle \mathbf{a}, \boldsymbol{\Sigma}^{1/2}\left(\mathbf{X}^{[k]}-\mathbf{Y}^{[k]}\right)\mathbf{b}\right\rangle = \sqrt{d_s}\Delta_{s\theta}^{[k]}\,\mathbf{a}(s)\mathbf{b}(\theta),$$

where

$$\Delta_{s\theta} := \mathbf{X}_{s\theta} - \mathbf{X}''_{s\theta}, \quad \text{and} \quad \Delta_{s\theta}^{[k]} := \begin{cases} \mathbf{X}'_{s\theta} - \mathbf{X}''_{s\theta} & \text{if } (s,\theta)\in S_k, \\ \mathbf{X}_{s\theta} - \mathbf{X}'''_{s\theta} & \text{if } (s,\theta)\notin S_k. \end{cases}$$

Therefore,

$$\sqrt{d_s}\Delta_{s\theta}\,\mathbf{f}(s)\mathbf{g}(\theta) \le \sigma - \mu \le \sqrt{d_s}\Delta_{s\theta}\mathbf{v}(s)\mathbf{u}(\theta),$$

and

$$\sqrt{d_s}\Delta_{s\theta}^{[k]}\,\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta) \le \sigma^{[k]} - \mu^{[k]} \le \sqrt{d_s}\Delta_{s\theta}^{[k]}\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta).$$

Consider

$$T_1 := d_s\left(\Delta_{s\theta}\mathbf{v}(s)\mathbf{u}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right), \quad T_2 := d_s\left(\Delta_{s\theta}\,\mathbf{f}(s)\mathbf{g}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\,\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta)\right),$$

$$T_3 := d_s\left(\Delta_{s\theta}\mathbf{v}(s)\mathbf{u}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\,\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta)\right), \quad T_4 := d_s\left(\Delta_{s\theta}\,\mathbf{f}(s)\mathbf{g}(\theta)\right)\left(\Delta_{s\theta}^{[k]}\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right).$$

Then we have

$$\min(T_1,T_2,T_3,T_4) \le (\sigma-\mu)\left(\sigma^{[k]}-\mu^{[k]}\right) \le \max(T_1,T_2,T_3,T_4). \tag{23}$$

To estimate (23), we introduce the following events

$$\mathcal{E}_1 := \left\{\max\left(\|\mathbf{v}\|_\infty, \|\mathbf{u}\|_\infty, \|\mathbf{f}\|_\infty, \|\mathbf{g}\|_\infty, \|\mathbf{v}^{[k]}\|_\infty, \|\mathbf{u}^{[k]}\|_\infty, \|\mathbf{f}^{[k]}\|_\infty, \|\mathbf{g}^{[k]}\|_\infty\right) \le n^{-\frac{1}{2}+\varepsilon}\right\}, \tag{24}$$

and

$$\mathcal{E}_2 := \left\{\max\left(\|\mathbf{v}-\mathbf{g}\|_\infty, \|\mathbf{u}-\mathbf{f}\|_\infty, \|\mathbf{v}^{[k]}-\mathbf{g}^{[k]}\|_\infty, \|\mathbf{u}^{[k]}-\mathbf{f}^{[k]}\|_\infty\right) \le n^{-\frac{1}{2}-\delta}\right\}. \tag{25}$$

Define the event $\mathcal{E} := \mathcal{E}_1\cap\mathcal{E}_2$. On the event $\mathcal{E}$, for all

$$J \in \left\{\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta), \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta), \mathbf{f}(s)\mathbf{g}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta), \mathbf{f}(s)\mathbf{g}(\theta)\mathbf{f}^{[k]}(s)\mathbf{g}^{[k]}(\theta)\right\}$$

we have

$$\left|J - \mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right| = O\left(n^{-2-\delta+3\varepsilon}\right). \tag{26}$$

Let $T := \min(T_1,T_2,T_3,T_4)$. On the event $\mathcal{E}$, using (26) we have

$$T \ge d_s\left(\Delta_{s\theta}\Delta_{s\theta}^{[k]}\right)\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta) - O\left(d_s\left|\Delta_{s\theta}\Delta_{s\theta}^{[k]}\right| n^{-2-\delta+3\varepsilon}\right). \tag{27}$$

Step 2. Next we claim that the contribution of T when $\mathcal{E}$ does not hold is negligible. Specifically, we have

$$\mathbb{E}\left[T\mathbb{1}_{\mathcal{E}^c}\right] = o(n^{-3}). \tag{28}$$

Recall that $d_s = \Theta(1)$. Without loss of generality, it suffices to show that

$$\mathbb{E}\left[\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\mathbb{1}_{\mathcal{E}^c}\right] = o(n^{-3}). \tag{29}$$

To see this, using $\mathbb{1}_{\mathcal{E}^c} = \mathbb{1}_{\mathcal{E}_1\backslash\mathcal{E}} + \mathbb{1}_{\mathcal{E}_1^c}$, we decompose the expectation into two parts

$$\mathbb{E}\left[\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\mathbb{1}_{\mathcal{E}^c}\right] = I_1 + I_2,$$

where

$$I_1 := \mathbb{E}\left[\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\mathbb{1}_{\mathcal{E}_1\backslash\mathcal{E}}\right], \quad I_2 := \mathbb{E}\left[\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\mathbb{1}_{\mathcal{E}_1^c}\right].$$

For the first term I_1, by delocalization and the relation $\mathcal{E}_1\backslash\mathcal{E} = \mathcal{E}_1\backslash\mathcal{E}_2$, we have

$$|I_1| \le n^{-2+4\varepsilon}\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_1\backslash\mathcal{E}_2}\right] \le n^{-2+4\varepsilon}\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_2^c}\right]. \tag{30}$$

Note that the random variable $\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}$ and the event $\mathcal{E}_2$ are dependent. To decouple these variables, we introduce a new event. Consider the event $\mathcal{E}_3 := \mathcal{A}\cup\mathcal{B}$, where

$$\mathcal{A} := \left\{\min\left(\sigma_1-\sigma_2, \sigma_1^{[k]}-\sigma_2^{[k]}\right) \ge n^{-1-c}\right\}, \quad \mathcal{B} := \left\{\min\left(\mu_1-\mu_2, \mu_1^{[k]}-\mu_2^{[k]}\right) \ge n^{-1-c}\right\}$$

Then,

$$\begin{aligned}
\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_3^c}\right] &\lesssim \mathbb{E}\left[\left(\Delta_{s\theta}^2 + (\Delta_{s\theta}^{[k]})^2\right)\mathbb{1}_{\mathcal{E}_3^c}\right] \\
&\lesssim \mathbb{E}\left[\left(\mathbf{X}_{s\theta}^2 + (\mathbf{X}'_{s\theta})^2 + (\mathbf{X}''_{s\theta})^2 + (\mathbf{X}'''_{s\theta})^2\right)\mathbb{1}_{\mathcal{E}_3^c}\right] \\
&\lesssim \mathbb{E}\left[\left(\mathbf{X}_{s\theta}^2 + (\mathbf{X}'_{s\theta})^2\right)\mathbb{1}_{\mathcal{B}^c}\right] + \mathbb{E}\left[\left((\mathbf{X}''_{s\theta})^2 + (\mathbf{X}'''_{s\theta})^2\right)\mathbb{1}_{\mathcal{A}^c}\right].
\end{aligned}$$

Observe that the random variables $\mathbf{X}_{s\theta}$ and $\mathbf{X}'_{s\theta}$ are independent of the event $\mathcal{B}$, and the random variable $\mathbf{X}''_{s\theta}$ is independent of $\mathcal{A}$. Therefore, by Lemma B.3,

$$\mathbb{E}\left[\left(\mathbf{X}_{s\theta}^2 + (\mathbf{X}'_{s\theta})^2\right)\mathbb{1}_{\mathcal{B}^c}\right] = O(n^{-1-\kappa}), \quad \mathbb{E}\left[\left((\mathbf{X}''_{s\theta})^2 + (\mathbf{X}'''_{s\theta})^2\right)\mathbb{1}_{\mathcal{A}^c}\right] = O(n^{-1-\kappa}).$$

By Lemma C.1, we have $\mathbb{P}(\mathcal{E}_3\backslash\mathcal{E}_2) = O(N^{-D})$ for any fixed large $D > 0$. Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_2^c}\right] &= \mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_3^c}\right] + \mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_3\backslash\mathcal{E}_2}\right] \\
&= O(n^{-1-\kappa}) + \sqrt{\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|^2\right]}\sqrt{\mathbb{P}(\mathcal{E}_3\backslash\mathcal{E}_2)} \\
&= O(n^{-1-\kappa}) + O(N^{-D}) \\
&= O(n^{-1-\kappa}).
\end{aligned}$$

Choosing $4\varepsilon < \kappa$, then (30) yields

$$|I_1| \le O(n^{-2+4\varepsilon-1-\kappa}) = o(n^{-3}).$$

For the term I_2, note that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}^{[k]}$ and $\mathbf{v}^{[k]}$ are unit vectors. We have that

$$\max(\|\mathbf{u}\|_\infty, \|\mathbf{v}\|_\infty, \|\mathbf{u}^{[k]}\|_\infty, \|\mathbf{v}^{[k]}\|_\infty) \le 1.$$

Recall that $\mathcal{E}_1$ holds with overwhelming probability. By the Cauchy-Schwarz inequality, for any large $D > 0$, we have

$$|I_2| \le \mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|\mathbb{1}_{\mathcal{E}_1^c}\right] \le \sqrt{\mathbb{E}\left[\left|\Delta_{s\theta}\,\Delta_{s\theta}^{[k]}\right|^2\right]}\sqrt{\mathbb{P}(\mathcal{E}_1^c)} = O(N^{-D}).$$

Hence we have shown the desired claim (29).

Step 3. Combining (23), (27) and (28), we obtain

$$\mathbb{E}\left[(\sigma-\mu)\left(\sigma^{[k]}-\mu^{[k]}\right)\right] \geq \mathbb{E}\left[d_s \Delta_{s\theta}\Delta_{s\theta}^{[k]}\,\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right] + o(n^{-3}).$$

Since $\frac{np+1}{np} \leq 2$ and $d_s = \Theta(1)$, by (21) we have

$$\mathbb{E}\left[\Delta_{s\theta}\Delta_{s\theta}^{[k]}\,\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right] \lesssim \frac{\mathsf{Var}(\sigma)}{k} + o(n^{-3}). \tag{31}$$

Since the random index (s,θ) is uniformly sampled, we have

$$\mathbb{E}\left[\Delta_{s\theta}\Delta_{s\theta}^{[k]}\,\mathbf{v}(s)\mathbf{u}(\theta)\mathbf{v}^{[k]}(s)\mathbf{u}^{[k]}(\theta)\right] = \frac{1}{np}\mathbb{E}\left[\sum_{1\leq i\leq n, 1\leq\alpha\leq p} \Delta_{i\alpha}\Delta_{i\alpha}^{[k]}\,\mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right]. \tag{32}$$

Note that

$$\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} = \begin{cases} (\mathbf{X}_{i\alpha}-\mathbf{X}'_{i\alpha})(\mathbf{X}'_{i\alpha}-\mathbf{X}''_{i\alpha}) & \text{if } (i,\alpha)\in S_k, \\ (\mathbf{X}_{i\alpha}-\mathbf{X}'_{i\alpha})(\mathbf{X}_{i\alpha}-\mathbf{X}'''_{i\alpha}) & \text{if } (i,\alpha)\notin S_k. \end{cases}$$

In either case, we have $\mathbb{E}[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]}] = p^{-1}$. Therefore,

$$\sum_{1\leq i\leq n, 1\leq\alpha\leq p} \mathbb{E}\left[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} \mid S_k\right]\mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha) = \frac{1}{p}\langle\mathbf{v},\mathbf{v}^{[k]}\rangle\langle\mathbf{u},\mathbf{u}^{[k]}\rangle.$$

Consequently, this implies

$$\mathbb{E}\left[\sum_{1\leq i\leq n, 1\leq\alpha\leq p} \mathbb{E}\left[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} \mid S_k\right]\mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right] = \frac{1}{p}\mathbb{E}\left[\langle\mathbf{v},\mathbf{v}^{[k]}\rangle\langle\mathbf{u},\mathbf{u}^{[k]}\rangle\right]. \tag{33}$$

Moreover, we claim that

$$\mathbb{E}\left[\sum_{1\leq i\leq n, 1\leq\alpha\leq p}\left(\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} - \mathbb{E}\left[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]} \mid S_k\right]\right)\mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right] = o(n^{-1}). \tag{34}$$

For the ease of notations, we set $\Xi_{i\alpha} := \Delta_{i\alpha}\Delta_{i\alpha}^{[k]} - \mathbb{E}[\Delta_{i\alpha}\Delta_{i\alpha}^{[k]}|S_k]$. It suffices to show that for all pairs (i,α) we have

$$\mathbb{E}\left[\Xi_{i\alpha}\mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha)\right] = o(n^{-3}). \tag{35}$$

To see this, we introduce another copy of $\mathbf{X}$, denoted by $\mathbf{X}''''$, which is independent of all previous random variables $(\mathbf{X},\mathbf{X}',\mathbf{X}'',\mathbf{X}''')$. For an arbitrarily fixed index (i,α), we define matrices $\widehat{\mathbf{X}}_{(i,\alpha)}$ and $\widehat{\mathbf{X}}_{(i,\alpha)}^{[k]}$ by resampling the (i,α) entry of $\mathbf{X}$ and $\mathbf{X}^{[k]}$ with $\mathbf{X}''''_{i\alpha}$. Let $\widehat{\mathbf{u}}$, $\widehat{\mathbf{v}}$ be the left and right top singular vector of $\widehat{\mathbf{X}}$, and similarly $\widehat{\mathbf{u}}^{[k]}$, $\widehat{\mathbf{v}}^{[k]}$ for $\widehat{\mathbf{X}}^{[k]}$. Define

$$\psi_{i\alpha} := \mathbf{v}(i)\mathbf{u}(\alpha)\mathbf{v}^{[k]}(i)\mathbf{u}^{[k]}(\alpha), \quad \widehat{\psi}_{i\alpha} := \widehat{\mathbf{v}}(i)\widehat{\mathbf{u}}(\alpha)\widehat{\mathbf{v}}^{[k]}(i)\widehat{\mathbf{u}}^{[k]}(\alpha).$$

A crucial observation is that $\Xi_{i\alpha}$ and $\widehat{\psi}_{i\alpha}$ are independent. This is because, conditioned on S_k, the matrices $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{X}}^{[k]}$ are independent of $(\mathbf{X}_{i\alpha},\mathbf{X}'_{i\alpha},\mathbf{X}''_{i\alpha},\mathbf{X}'''_{i\alpha})$. Such a conditional independence is also true for the singular vectors, and hence also holds for $\widehat{\psi}_{i\alpha}$. On the other hand, by definition, the variable $\Xi_{i\alpha}$ only depends on $(\mathbf{X}_{i\alpha}, \mathbf{X}'_{i\alpha}, \mathbf{X}''_{i\alpha}, \mathbf{X}'''_{i\alpha})$. Therefore,

$$\mathbb{E}\left[\Xi_{i\alpha}\widehat{\psi}_{i\alpha}\right] = \mathbb{E}\left[\mathbb{E}[\Xi_{i\alpha}|S_k]\,\mathbb{E}[\widehat{\psi}_{i\alpha}|S_k]\right] = 0$$

Thus, we reduce (35) to showing

$$\mathbb{E}\left[\Xi_{i\alpha}\left(\psi_{i\alpha} - \widehat{\psi}_{i\alpha}\right)\right] = o(n^{-3}). \tag{36}$$

The proof of (36) is similar as previous arguments. Consider the events

$$\widehat{\mathcal{E}}_1 := \left\{\max\left(\|\mathbf{v}\|_\infty, \|\mathbf{u}\|_\infty, \|\widehat{\mathbf{u}}\|_\infty, \|\widehat{\mathbf{v}}\|_\infty, \|\mathbf{v}^{[k]}\|_\infty, \|\mathbf{u}^{[k]}\|_\infty, \|\widehat{\mathbf{u}}^{[k]}\|_\infty, \|\widehat{\mathbf{v}}^{[k]}\|_\infty\right) \le n^{-\frac{1}{2}+\varepsilon}\right\},$$

$$\widehat{\mathcal{E}}_2 := \left\{\max\left(\|\mathbf{v} - \widehat{\mathbf{v}}\|_\infty, \|\mathbf{u} - \widehat{\mathbf{u}}\|_\infty, \|\mathbf{v}^{[k]} - \widehat{\mathbf{v}}^{[k]}\|_\infty, \|\mathbf{u}^{[k]} - \widehat{\mathbf{u}}^{[k]}\|_\infty\right) \le n^{-\frac{1}{2}-\delta}\right\}.$$

On the event $\widehat{\mathcal{E}} := \widehat{\mathcal{E}}_1 \cap \widehat{\mathcal{E}}_2$, we have $|\psi_{i\alpha} - \widehat{\psi}_{i\alpha}| = O(n^{-2-\delta+3\varepsilon})$. Note that $\mathbb{E}[|\Xi_{i\alpha}|] = O(n^{-1})$ since $\mathbb{E}[|\Delta_{i\alpha}\Delta^{[k]}_{i\alpha}|] = O(n^{-1})$. As a consequence,

$$\mathbb{E}\left[\left|\Xi_{i\alpha}(\psi_{i\alpha} - \widehat{\psi}_{i\alpha})\right|\mathbb{1}_{\widehat{\mathcal{E}}}\right] = O(n^{-3-\delta+3\varepsilon}) = o(n^{-3}). \tag{37}$$

Using the same argument as in (29), we have

$$\mathbb{E}\left[\left|\Xi_{i\alpha}(\psi_{i\alpha} - \widehat{\psi}_{i\alpha})\right|\mathbb{1}_{\widehat{\mathcal{E}}^c}\right] \lesssim N^{-2+4\varepsilon}\mathbb{E}\left[|\Xi_{i\alpha}|\mathbb{1}_{\widehat{\mathcal{E}}^c}\right] = O(N^{-2+4\varepsilon}N^{-1-\kappa}) = o(n^{-3}), \tag{38}$$

where κ is the constant in the gap property (Lemma B.3) Thus, by (37) and (38), we have shown the desired claim (36).

Based on (31) and (32), combining (33) and (34) yields

$$\frac{1}{np^2}\mathbb{E}\left[\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle\right] \lesssim \frac{\mathsf{Var}(\sigma)}{k} + o\left(\frac{1}{n^3}\right) + o\left(\frac{1}{n^2 p}\right)$$

By Lemma B.2 we have $\mathsf{Var}(\sigma) = O(n^{-4/3+\varepsilon_0/2})$ and the assumption $k \ge n^{5/3+\varepsilon_0}$, we have

$$\mathbb{E}\left[\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle\right] \le \frac{np^2}{k}O(n^{-4/3+\varepsilon_0/2}) + o(1) = o(1).$$

This implies

$$\mathbb{E}\left[|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle|\right] \to 0. \tag{39}$$

In the null model, we have $\mathsf{Var}(\sigma) = O(n^{-4/3})$, and therefore the threshold can be improved to $k \gg n^{5/3}$.

Step 4. Consider the symmetrization matrix $\widetilde{\mathbf{X}}$ defined in (6). The variational representation of the top eigenvalue yields

$$\sigma = \frac{\langle \mathbf{w}, \widetilde{\mathbf{X}}\mathbf{w}\rangle}{\|\mathbf{w}\|_2^2}, \quad \sigma^{[k]} = \frac{\langle \mathbf{w}^{[k]}, \widetilde{\mathbf{X}}^{[k]}\mathbf{w}^{[k]}\rangle}{\|\mathbf{w}^{[k]}\|_2^2} \quad \text{with } \|\mathbf{w}\|_2^2 = \|\mathbf{w}^{[k]}\|_2^2 = 2.$$

Using the same arguments as in Step 1-3, we can conclude that

$$\mathbb{E}\left[|\langle \mathbf{w}, \mathbf{w}^{[k]}\rangle|^2\right] = \mathbb{E}\left[|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle + \langle \mathbf{u}, \mathbf{u}^{[k]}\rangle|^2\right] \to 0.$$

Combined with (39), this gives us

$$\mathbb{E}\left[|\langle \mathbf{v}, \mathbf{v}^{[k]}\rangle|^2 + |\langle \mathbf{u}, \mathbf{u}^{[k]}\rangle|^2\right] \to 0,$$

which proves the desired results.

D Proofs for the Stability Regime for the Weakly Spiked Model

Throughout the whole section, we will focus on the behaviour of $\mathbf{v}$ and $\mathbf{v}^{[k]}$. Similar results also hold for $\mathbf{u}$ and $\mathbf{u}^{[k]}$ via the same arguments.

D.1 Linearization and local law of resolvent

In the study of sample covariance matrices, a convenient trick is to consider the symmetrization $\widetilde{\mathbf{X}}$ of the data matrix $\mathbf{\Sigma}^{1/2}\mathbf{X}$ (defined as in (6)) when exploring its spectral properties. For $z \in \mathbb{C}$ with $\operatorname{Im} z > 0$, We introduce the resolvent of this symmetrization

$$\mathbf{R}(z) := \begin{pmatrix} -\mathbf{I}_n & (\mathbf{\Sigma}^{1/2}\mathbf{X})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}) & -z\mathbf{I}_p \end{pmatrix}^{-1}. \tag{40}$$

Note that $\mathbf{R}(z)$ is not the conventional definition of the resolvent matrix, but we still call it resolvent for convenience. For the ease of notations, we will relabel the indices in $\mathbf{R}$ in the following way:

Definition D.1 (Index sets)**.** We define the index sets

$$\mathcal{I}_1 := \{1, \dots, n\}, \quad \mathcal{I}_2 := \{1, \dots, p\}, \quad \mathcal{I} := \mathcal{I}_1 \cup \{n + i : i \in \mathcal{I}_2\}.$$

For a general matrix $\mathbf{M} \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$, we label the indices of the matrix elements in the following way: for $a, b \in \mathcal{I}$, if $1 \le a, b \le n$, then typically we use Greek letters α, β to represent them; if $n + 1 \le a, b \le n + p$, we use the corresponding Latin letters $i = a - n$ and $j = b - n$ to represent them.

The resolvent $\mathbf{R}$ is closely related to the eigenvalue and eigenvectors of the sample covariance matrix. As discussed in [DY18][Equation (3.9),(3.10)], we have

$$\mathbf{R}_{\alpha\beta}(z) = \sum_{\ell=1}^{n} \frac{z\mathbf{u}_\ell(\alpha)\mathbf{u}_\ell(\beta)}{\lambda_\ell - z}, \quad \mathbf{R}_{ij}(z) = \sum_{\ell=1}^{p} \frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{\lambda_\ell - z}, \tag{41}$$

and

$$\mathbf{R}_{i\alpha}(z) = \sum_{\ell=1}^{p} \frac{\sqrt{\lambda_\ell}\mathbf{u}_\ell(\alpha)\mathbf{v}_\ell(i)}{\lambda_\ell - z}, \quad \mathbf{R}_{\alpha i}(z) = \sum_{\ell=1}^{p} \frac{\sqrt{\lambda_\ell}\mathbf{v}_\ell(i)\mathbf{u}_\ell(\alpha)}{\lambda_\ell - z}.$$

An important result is the local deformed Marchenko-Pastur law for the resolvent matrix $\mathbf{R}$. This was first proved in [BKYY16], and we refer to [DY18][Lemma 3.11] for a version that is consistent with our setting. Specifically, the resolvent matrix $\mathbf{R}$ has a deterministic limit, defined by

$$\mathbf{G}(z) := \begin{pmatrix} -(1 + m_{\mathsf{fc}}(z)\mathbf{\Sigma})^{-1} & 0 \\ 0 & m_{\mathsf{fc}}(z)\mathbf{I}_p \end{pmatrix}, \tag{42}$$

where $m_{\mathsf{fc}}(z)$ is the Stieltjes transform of the deformed Marchenko-Pasture law given by (9)

To state the local law, we will focus on the spectral domain

$$\mathcal{S} := \{E + \mathrm{i}\eta : \lambda_{\mathrm{R}} - 1 \le E \le \lambda_{\mathrm{R}} + 1, 0 < \eta < 1\}. \tag{43}$$

Lemma D.2 (Local deformed Marchenko-Pastur law)**.** *For any $\varepsilon > 0$, the following estimate holds With overwhelming probability uniformly for $z \in \mathcal{S}$,*

$$\max_{a,b\in\mathcal{I}} |\mathbf{R}_{ab}(z) - \mathbf{G}_{ab}(z)| \le n^{\varepsilon} \left(\sqrt{\frac{\operatorname{Im} m_{\mathsf{fc}}(z)}{n\eta}} + \frac{1}{n\eta} \right). \tag{44}$$

To give a precise characterization of the resolvent, we rely on the following estimates for the Stieltjes transform $m_{\mathsf{MP}}(z)$ of the Marchenko-Pasture law. We refer to e.g. [BKYY16][Lemma 3.6] and [DY18][Lemma 3.6] for more details.

Lemma D.3 (Estimate for $m_{\mathsf{fc}}(z)$)**.** *For $z = E + \mathrm{i}\eta$, let $\kappa(z) := \min(|E - \lambda_{\mathrm{L}}|, |E - \lambda_{\mathrm{R}}|)$ denote the distance to the spectral edge. If $z \in \mathcal{S}$, we have*

$$|m_{\mathsf{fc}}(z)| \asymp 1, \quad \textit{and} \quad \operatorname{Im} m_{\mathsf{fc}}(z) \asymp \begin{cases} \sqrt{\kappa(z) + \eta} & \textit{if } E \in [\lambda_{\mathrm{L}}, \lambda_{\mathrm{R}}], \\ \frac{\eta}{\sqrt{\kappa(z)+\eta}} & \textit{if } E \notin [\lambda_{\mathrm{L}}, \lambda_{\mathrm{R}}]. \end{cases} \tag{45}$$

In the following analysis, we will work with $z = E + \mathrm{i}\eta$ satisfying $|E - \lambda_{\mathrm{R}}| \leq n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$, where $0 < \delta < \frac{1}{3}$ is some parameter. Uniformly in this regime, the local law yields that the following is true with overwhelming probability for all $\varepsilon > 0$ and some universal constant $C_0 > 0$,

$$\sup_z \max_{a \neq b \in \mathcal{I}} |\mathbf{R}_{ab}(z)| \leq n^{-\frac{1}{3}+\delta+\varepsilon}, \quad \text{and} \quad \sup_z \max_{a \in \mathcal{I}} |\mathbf{R}_{aa}(z)| \leq C_0. \tag{46}$$

These estimates will be used repeatedly in the following subsections.

D.2 Stability of the resolvent

In this subsection, we will prove the main technical result for the proof of resampling stability in the weakly spiked model. Specifically, we will show that under moderate resampling, the resolvent matrices are stable. Since resolvent is closely related to various spectral statistics, this stability lemma for resolvent will be a key ingredient for our proof.

Lemma D.4. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. Assume $k \leq n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$. There exists $\delta_0 > 0$ such that for all $0 < \delta < \delta_0$, uniformly for $z = E + \mathrm{i}\eta$ with $|E - \lambda_{\mathrm{R}}| \leq n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$, there exists $c > 0$ such that the following is true with overwhelming probability*

$$\max_{i,j} \left| \mathbf{R}_{ij}^{[k]}(z) - \mathbf{R}_{ij}(z) \right| \leq \frac{1}{n^{1+c}\eta}, \quad \max_{\alpha,\beta} \left| \mathbf{R}_{\alpha\beta}^{[k]}(z) - \mathbf{R}_{\alpha\beta}(z) \right| \leq \frac{1}{n^{1+c}\eta}. \tag{47}$$

Proof. Recall that $S_k := \{(i_1, \alpha_1), \ldots, (i_k, \alpha_k)\}$ is the random subset of matrix indices whose elements are resampled in the matrix $\mathbf{X}$. For $1 \leq t \leq k$, let $\mathbf{X}^{[t]}$ be the matrix obtained from $\mathbf{X}$ by resampling the $\{(i_s, \alpha_s)\}_{1 \leq s \leq t}$ entries and let $\mathcal{F}_t$ be the σ-algebra generated by the random variables $\mathbf{X}$, S_k and $\{\mathbf{X}'_{i_s\alpha_s}\}_{1 \leq s \leq t}$. For $a, b \in \mathcal{I}$, we define

$$T_{ab} := \{t : \{i_t, \alpha_t\} \cap \{a, b\} \neq \emptyset\}.$$

Let $\varepsilon > 0$ be an arbitrarily fixed parameter, and let C_0 be the constant in (46). Consider the event $\mathcal{E}_t \in \mathcal{F}_t$ where for all $z = E + \mathrm{i}\eta$ with $|z - \lambda_{\mathrm{R}}| \leq n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$ we have

$$\max_{a \neq b} \left| \mathbf{R}_{ab}^{[t]}(z) \right| \leq n^{-\frac{1}{3}+\delta+\varepsilon} =: \Psi, \quad \text{and} \quad \max_a \left| \mathbf{R}_{aa}^{[t]}(z) \right| \leq C_0.$$

Define $\mathbf{X}_0^{[t]}$ as the matrix obtained from $\mathbf{X}^{[t]}$ by replacing the (i_t, α_t) entry with 0, and also define its symmetrization $\widetilde{\mathbf{X}}_0^{[t]} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{I}|}$ as in (6). Note that $\widetilde{\mathbf{X}}_0^{[t+1]}$ is $\mathcal{F}_t$-measurable. We write

$$\widetilde{\mathbf{X}}^{[t]} = \widetilde{\mathbf{X}}_0^{[t+1]} + \widetilde{\mathbf{P}}^{[t+1]}, \quad \widetilde{\mathbf{X}}^{[t+1]} = \widetilde{\mathbf{X}}_0^{[t+1]} + \widetilde{\mathbf{Q}}^{[t+1]},$$

where $\widetilde{\mathbf{P}}^{[t]}, \widetilde{\mathbf{Q}}^{[t]}$ are $|\mathcal{I}| \times |\mathcal{I}|$ symmetric matrices whose elements are all 0 except at the (i_t, α_t) and (α_t, i_t) entries, satisfying

$$(\widetilde{\mathbf{P}}^{[t]})_{ab} = \begin{cases} \sqrt{d_{i_t}}\mathbf{X}_{i_t\alpha_t} & \text{if } \{a,b\} = \{i_t, \alpha_t\}, \\ 0 & \text{otherwise} \end{cases} \quad (\widetilde{\mathbf{Q}}^{[t]})_{ab} = \begin{cases} \sqrt{d_{i_t}}\mathbf{X}'_{i_t\alpha_t} & \text{if } \{a,b\} = \{i_t, \alpha_t\}, \\ 0 & \text{otherwise} \end{cases}.$$

Define the resolvents for the matrices $\widetilde{\mathbf{X}}^{[t]}$ and $\widetilde{\mathbf{X}}_0^{[t]}$ as in (40):

$$\mathbf{R}^{[t]} := \begin{pmatrix} -\mathbf{I}_n & (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[t]})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}^{[t]}) & -z\mathbf{I}_p \end{pmatrix}^{-1}, \quad \mathbf{R}_0^{[t]} := \begin{pmatrix} -\mathbf{I}_n & (\mathbf{\Sigma}^{1/2}\mathbf{X}_0^{[t]})^\top \\ (\mathbf{\Sigma}^{1/2}\mathbf{X}_0^{[t]}) & -z\mathbf{I}_p \end{pmatrix}^{-1}.$$

Using first-order resolvent expansion, we obtain

$$\mathbf{R}_0^{[t+1]} = \mathbf{R}^{[t]} + \mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]} + \left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\right)^2 \mathbf{R}_0^{[t+1]}. \tag{48}$$

The triangle inequality yields

$$\left|\left(\mathbf{R}_0^{[t+1]} - \mathbf{R}^{[t]}\right)_{ij}\right| \le \left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\right)_{ij}\right| + \left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]}\right)_{ij}\right|.$$

Note that $\widetilde{\mathbf{P}}^{[t+1]}$ has only two non-zero entries,

$$\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\right)_{ij} = \sum_{\ell_1,\ell_2} \mathbf{R}^{[t]}_{i\ell_1}\widetilde{\mathbf{P}}^{[t+1]}_{\ell_1\ell_2}\mathbf{R}^{[t]}_{\ell_2 j} = \sqrt{d_{i_{t+1}}}X_{i_{t+1}\alpha_{t+1}}\left(\mathbf{R}^{[t]}_{ii_{t+1}}\mathbf{R}^{[t]}_{\alpha_{t+1}j} + \mathbf{R}^{[t]}_{i\alpha_{t+1}}\mathbf{R}^{[t]}_{i_{t+1}j}\right)$$

Recall that and $|X_{i_{t+1}\alpha_{t+1}}| \le n^{-1/2+\varepsilon}$ with overwhelming probability thanks to the sub-exponential decay (see Assumption 1). Then on the event $\mathcal{E}_t$, we have

$$\left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\right)_{ij}\right| \le 2\sqrt{d_{i_{t+1}}}C_0\Psi n^{-\frac{1}{2}+\varepsilon}.$$

Similarly,

$$\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]}\right)_{ij} = \sum_{\{m_1,m_2\},\{m_3,m_4\}=\{i_{t+1},\alpha_{t+1}\}} \mathbf{R}^{[t]}_{im_1}\widetilde{\mathbf{P}}^{[t+1]}_{m_1m_2}\mathbf{R}^{[t]}_{m_2m_3}\widetilde{\mathbf{P}}^{[t+1]}_{m_3m_4}(\mathbf{R}_0^{[t+1]})_{m_4j}.$$

We use the trivial bound $|\mathbf{R}_0^{[t+1]}| \le \eta^{-1}$ for the last term. Then, on the event $\mathcal{E}_t$, we have

$$\left|\left(\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}^{[t]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]}\right)_{ij}\right| \le 2d_{i_{t+1}}n^{-1+2\varepsilon}\eta^{-1}(\Psi^2 + C_0^2) \ll \Psi.$$

Therefore, we have shown that, on the event $\mathcal{E}_t$,

$$\max_{i\ne j}\left|(\mathbf{R}_0^{[t+1]})_{ij}\right| \le 2\Psi, \quad \max_i \left|(\mathbf{R}_0^{[t+1]})_{ii}\right| \le 2C_0. \tag{49}$$

Similarly, using the first-order resolvent expansion for $\mathbf{R}^{[t+1]}$ around $\mathbf{R}^{[t]}$, we have

$$\mathbf{R}^{[t+1]} = \mathbf{R}^{[t]} + \mathbf{R}^{[t]}(\widetilde{\mathbf{P}}^{[t+1]} - \widetilde{\mathbf{Q}}^{[t+1]})\mathbf{R}^{[t]} + \left(\mathbf{R}^{[t]}(\widetilde{\mathbf{P}}^{[t+1]} - \widetilde{\mathbf{Q}}^{[t+1]})\right)^2 \mathbf{R}^{[t+1]}.$$

By the same arguments as above, on the event $\mathcal{E}_t$, we can derive

$$\max_{i\neq j}\left|\mathbf{R}_{ij}^{[t+1]}\right|\leq 2\Psi,\quad \max_{i}\left|\mathbf{R}_{ii}^{[t+1]}\right|\leq 2C_0.$$

Next, we use the resolvent identity (or zeroth-order expansion) for $\mathbf{R}^{[t+1]}$ and $\mathbf{R}_0^{[t+1]}$:

$$\mathbf{R}^{[t+1]}=\mathbf{R}_0^{[t+1]}-\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\mathbf{R}^{[t+1]}.$$

This leads to

$$\left|\left(\mathbf{R}^{[t+1]}-\mathbf{R}_0^{[t+1]}\right)_{ij}\right|=\left|\sum_{\{\ell_1,\ell_2\}=\{i_{t+1}\alpha_{t+1}\}}(\mathbf{R}_0^{[t+1]})_{i\ell_1}\widetilde{\mathbf{Q}}_{\ell_1\ell_2}^{[t+1]}\mathbf{R}_{\ell_2 j}^{[t+1]}\right|$$

Thus, on the event $\mathcal{E}_t$, we conclude

$$\left|\left(\mathbf{R}^{[t+1]}-\mathbf{R}_0^{[t+1]}\right)_{ij}\right|\leq 4\sqrt{d_{i_{t+1}}}n^{-\frac{1}{2}+\varepsilon}\left(\Psi^2+C_0\Psi\mathbb{1}_{((t+1)\in T_{ij})}\right)=:\mathfrak{f}_{ij}^{[t+1]}\tag{50}$$

Meanwhile, the second-order resolvent expansion of $\mathbf{R}^{[t+1]}$ around $\mathbf{R}_0^{[t+1]}$ yields

$$\mathbf{R}^{[t+1]}=\mathbf{R}_0^{[t+1]}-\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\mathbf{R}_0^{[t+1]}+\left(\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\right)^2\mathbf{R}_0^{[t+1]}-\left(\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{Q}}^{[t+1]}\right)^3\mathbf{R}^{[t+1]}.$$

A key observation is that $\mathbf{R}_0^{[t+1]}$ is $\mathcal{F}_t$-measurable, and $\mathbb{E}[\widetilde{\mathbf{Q}}^{[t+1]}|\mathcal{F}_t]=0$. For simplicity of notations, we set

$$\mathfrak{q}_{ij}^{[t]}:=\left((\mathbf{R}_0^{[t]}\widetilde{\mathbf{E}}^{(i_t,\alpha_t)})^2\mathbf{R}_0^{[t]}\right)_{ij}$$

where $\widetilde{\mathbf{E}}^{(i_t,\alpha_t)}\in\mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ is the symmetrization of the matrix $\mathbf{E}^{(i_t,\alpha_t)}\in\mathbb{R}^{p\times n}$ whose elements are all 0 except $\widetilde{\mathbf{E}}_{i_t\alpha_t}^{(i_t,\alpha_t)}=\widetilde{\mathbf{E}}_{\alpha_t i_t}^{(i_t,\alpha_t)}=1$. Then we have

$$\left|\mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]-(\mathbf{R}_0^{[t+1]})_{ij}-p^{-1}\mathfrak{q}_{ij}^{[t+1]}\right|\leq 32d_{i_{t+1}}^{\frac{3}{2}}n^{-\frac{3}{2}+3\varepsilon}\left(\Psi^2C_0^2+C_0^4\mathbb{1}_{((t+1)\in T_{ij})}\right)=:\mathfrak{g}_{ij}^{[t+1]}.\tag{51}$$

Similarly, using resolvent expansion of $\mathbf{R}^{[t]}$ around $\mathbf{R}_0^{[t+1]}$, we obtain

$$\mathbf{R}^{[t]}=\mathbf{R}_0^{[t+1]}-\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{P}}^{[t+1]}\mathbf{R}_0^{[t+1]}+(\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{P}}^{[t+1]})^2\mathbf{R}_0^{[t+1]}-(\mathbf{R}_0^{[t+1]}\widetilde{\mathbf{P}}^{[t+1]})^3\mathbf{R}^{[t]}.$$

By the same arguments as above, on the event $\mathcal{E}_t$, we deduce that

$$\left|\mathbf{R}_{ij}^{[t]}-(\mathbf{R}_0^{[t+1]})_{ij}+\mathbf{X}_{i_{t+1}\alpha_{t+1}}\mathfrak{p}_{ij}^{[t+1]}-\mathbf{X}_{i_{t+1}\alpha_{t+1}}^2\mathfrak{q}_{ij}^{[t+1]}\right|\leq\mathfrak{g}_{ij}^{[t+1]}\tag{52}$$

where

$$\mathfrak{p}_{ij}^{[t]}:=\left(\mathbf{R}_0^{[t]}\widetilde{\mathbf{E}}^{(i_t,\alpha_t)}\mathbf{R}_0^{[t]}\right)_{ij}.\tag{53}$$

Combining (51) and (52) yields

$$\left|\mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]-\mathbf{R}_{ij}^{[t]}-\mathbf{X}_{i_{t+1}\alpha_{t+1}}\mathfrak{p}_{ij}^{[t+1]}+(\mathbf{X}_{i_{t+1}\alpha_{t+1}}^2-p^{-1})\mathfrak{q}_{ij}^{[t+1]}\right|\leq 2\mathfrak{g}_{ij}^{[t+1]}.\tag{54}$$

By a telescopic summation, we obtain

$$\begin{aligned}\mathbf{R}_{ij}^{[k]}-\mathbf{R}_{ij}&=\sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]}-\mathbf{R}_{ij}^{[t]}\right)\\&=\sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]}-\mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right)+\sum_{t=0}^{k-1}\mathbf{X}_{i_{t+1}\alpha_{t+1}}\mathfrak{p}_{ij}^{[t+1]}+\sum_{t=0}^{k-1}(\mathbf{X}_{i_{t+1}\alpha_{t+1}}^2-p^{-1})\mathfrak{q}_{ij}^{[t+1]}+\mathfrak{r}_{ij}\end{aligned}\tag{55}$$

where the remainder $\mathfrak{r}_{ij}$ is bounded by (54)

$$|\mathfrak{r}_{ij}| \leq 2\sum_{t=0}^{k-1} \mathfrak{g}_{ij}^{[t+1]}.$$

Recall the expression of $\mathfrak{g}_{ij}^{[t]}$, to estimate the remainder, we need to control the size of the set T_{ij}. Note that $\mathbb{E}\left[|T_{ij}|\right] = 2k/p$. By a Berstein-type inequality (see e.g. [Cha07][Proposition 1.1]), for any $x > 0$, we have

$$\mathbb{P}\left(|T_{ij}| \geq \mathbb{E}\left[|T_{ij}|\right] + x\right) \leq \exp\left(-\frac{x^2}{4\mathbb{E}\left[|T_{ij}|\right] + 2x}\right)$$

Recall that $k \leq n^{5/3-\epsilon_0}$. The inequality together with a union bound implies that

$$\max_{i,j} |T_{ij}| \leq \frac{3\max(k, p(\log n)^2)}{p} =: \mathsf{T}$$

with overwhelming probability. We denote this event by $\mathcal{T}$. On the event $\mathcal{T}$, we have

$$|\mathfrak{r}_{ij}| \lesssim 2kn^{-\frac{3}{2}+3\varepsilon}\Psi^2 C_0^2 + 2n^{-\frac{3}{2}+3\varepsilon} C_0^4 \mathsf{T} \lesssim n^{3\varepsilon}\sqrt{\mathsf{T}}\Psi^2. \tag{56}$$

For the first term in (55), we set

$$\mathfrak{w}_{ij}^{[t+1]} := \left(\mathbf{R}_{ij}^{[t+1]} - \mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right) \mathbb{1}_{\mathcal{E}_t}.$$

Note that $\mathcal{E}_t \in \mathcal{F}_t$. This implies that $\mathbb{E}[\mathfrak{w}_{ij}^{[t+1]}|\mathcal{F}_t] = 0$. Moreover, by (50), on the event $\mathcal{E}_t$ we have $|\mathfrak{w}_{ij}^{[t+1]}| \leq 2\mathfrak{f}_{ij}^{[t+1]}$. Further, on the event $\mathcal{T}$,

$$\left(\sum_{t=0}^{k-1} (\mathfrak{f}_{ij}^{[t+1]})^2\right)^{1/2} \lesssim n^{-\frac{1}{2}+\varepsilon}\Psi^2\sqrt{k} + n^{-\frac{1}{2}+\varepsilon} C_0 \Psi\sqrt{\mathsf{T}} \leq 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}.$$

Using the Azuma-Hoeffding inequality, for any $x \geq 0$, we have

$$\mathbb{P}\left(\left|\sum_{t=0}^{k-1} \mathfrak{w}_{ij}^{[t+1]}\right| \geq 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}x\right) \leq 2\exp\left(-\frac{x^2}{2}\right).$$

Moreover,

$$\mathbb{P}\left(\left|\sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]} - \mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right)\right| \geq 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}x\right) \leq \mathbb{P}\left(\left|\sum_{t=0}^{k-1}\mathfrak{w}_{ij}^{[t+1]}\right| \geq 2n^{\varepsilon}\Psi^2\sqrt{\mathsf{T}}x\right) + \sum_{t=0}^{k-1}\mathbb{P}(\mathcal{E}_t^c).$$

Recall that $\mathcal{E}_t$ holds with overwhelming probability, and consequently $\sum_{t=0}^{k-1}\mathbb{P}(\mathcal{E}_t^c) \leq n^{-D}$ for any $D > 0$. Choosing $x = n^{\varepsilon}$ implies that with overwhelming probability

$$\left|\sum_{t=0}^{k-1}\left(\mathbf{R}_{ij}^{[t+1]} - \mathbb{E}\left[\mathbf{R}_{ij}^{[t+1]}|\mathcal{F}_t\right]\right)\right| \leq 2n^{2\varepsilon}\Psi^2\sqrt{\mathsf{T}}. \tag{57}$$

For the next two terms in (55), we will deal with them by introducing a backward filtration. Let $\mathcal{F}_t'$ be the σ-algebra generated by the random variables $\mathbf{X}'$, S_k and $\{\mathbf{X}_{i\alpha}\}$ with $i \notin \{i_1, \ldots, i_t\}$

and $\alpha \notin \{\alpha_1, \ldots, \alpha_t\}$. Similarly as above, we consider the event $\mathcal{E}'_t$ that for all $z = E + \mathrm{i}\eta$ with $|z - \lambda_{\mathrm{R}}| \le n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$ we have

$$\max_{a \neq b} \left| \mathbf{R}^{[t]}_{ab}(z) \right| \le \Psi, \quad \text{and} \quad \max_{a} \left| \mathbf{R}^{[t]}_{aa}(z) \right| \le C_0.$$

Using resolvent expansion, the same arguments for (49) yield that, on the event $\mathcal{E}'_t$, we have

$$\max_{i \neq j} \left| (\mathbf{R}^{[t]}_0)_{ij} \right| \le 2\Psi, \quad \max_{i} \left| (\mathbf{R}^{[t]}_0)_{ii} \right| \le 2C_0.$$

A key observation is that $\mathfrak{p}^{[t]}_{ij}$ defined in (53) is $\mathcal{F}'_t$-measurable. Also, we have $\mathbb{E}[\mathbf{X}_{i_t\alpha_t} | \mathcal{F}'_t] = 0$. Consider

$$\widetilde{\mathfrak{p}}^{[t]}_{ij} := \mathbf{X}_{i_t\alpha_t} \mathfrak{p}^{[t]}_{ij} \mathbb{1}_{\mathcal{E}'_t}.$$

Then we have $\mathbb{E}[\widetilde{\mathfrak{p}}^{[t]}_{ij} | \mathcal{F}'_t] = 0$ since we also have $\mathcal{E}'_t \in \mathcal{F}'_t$. Note that

$$\mathbb{P}\left(\left| \sum_{t=0}^{k-1} \mathbf{X}_{i_{t+1}\alpha_{t+1}} \mathfrak{p}^{[t+1]}_{ij} \right| \ge x \right) \le \mathbb{P}\left(\left| \sum_{t=0}^{k-1} \widetilde{\mathfrak{p}}^{[t+1]}_{ij} \right| \ge x \right) + \sum_{t=0}^{k-1} \mathbb{P}((\mathcal{E}'_{t+1})^c),$$

The second term is negligible since $\mathcal{E}'_t$ holds with overwhelming probability. To estimate the first term, we use Azuma-Hoeffding inequality as before. Based on similar arguments as in (50), we deduce

$$\left| \widetilde{\mathfrak{p}}^{[t]}_{ij} \right| \le 4\sqrt{d_{i_t}} n^{-\frac{1}{2}+\varepsilon} \left(\Psi^2 + C_0 \Psi \mathbb{1}_{(t \in T_{\alpha\beta})} \right).$$

By considering the event $\mathcal{T}$ and using Azuma-Hoeffding inequality as in (57), we can conclude that with overwhelming probability,

$$\left| \sum_{t=0}^{k-1} \widetilde{\mathfrak{p}}^{[t+1]}_{ij} \right| \le n^{2\varepsilon} \Psi^2 \sqrt{\mathsf{T}}$$

As a consequence, with overwhelming probability

$$\left| \sum_{t=0}^{k-1} \mathbf{X}_{i_{t+1}\alpha_{t+1}} \mathfrak{p}^{[t+1]}_{ij} \right| \lesssim n^{2\varepsilon} \Psi^2 \sqrt{\mathsf{T}}. \tag{58}$$

For the third term in (55), by the same arguments, we have

$$\left| \sum_{t=0}^{k-1} (\mathbf{X}^2_{i_{t+1}\alpha_{t+1}} - p^{-1}) \mathfrak{q}^{[t+1]}_{ij} \right| \lesssim n^{2\varepsilon} \Psi^2 \sqrt{\mathsf{T}}. \tag{59}$$

Finally, combining (55), (56), (57), (58) and (59), we have shown that

$$\left| \mathbf{R}^{[k]}_{ij}(z) - \mathbf{R}_{ij}(z) \right| \lesssim n^{3\varepsilon} \Psi^2 \sqrt{\mathsf{T}}.$$

Recall that $\eta = n^{-2/3-\delta}$, $\Psi = O(n^{-\frac{1}{3}+\delta+\varepsilon})$, and $\mathsf{T} = O(n^{\frac{2}{3}-\epsilon_0})$. Then we obtain

$$n\eta \left| \mathbf{R}^{[k]}_{ij}(z) - \mathbf{R}_{ij}(z) \right| \le n^{-\frac{\epsilon_0}{2}+\delta+5\varepsilon}. \tag{60}$$

Choosing $\delta + 5\varepsilon < \frac{\epsilon_0}{2}$ yields the desired bound (47) for a fixed z.

So far, we have proved the desired result for a fixed z. To extend this result to a uniform estimate, we simply invoke a standard net argument. To do this, we divide the interval $[-n^{-2/3+\delta}, n^{-2/3+\delta}]$ into n^2 sub-intervals, and consider $z = E + \mathrm{i}\eta$ with $\kappa(z)$ taking values in each sub-interval. Note that

$$|\mathbf{R}_{ij}(z_1) - \mathbf{R}_{ij}(z_2)| \le \frac{|z_1 - z_2|}{\min(\mathrm{Im}\,(z_1), \mathrm{Im}\,(z_2))^2}.$$

For z_1, z_2 associated with the same sub-interval, we have

$$n\eta|\mathbf{R}_{ij}(z_1) - \mathbf{R}_{ij}(z_2)| \le n\eta \frac{n^{-2/3+\delta} n^{-2}}{\eta^2} \le n^{-1+2\delta},$$

which is of lower order compared with the error bound in (60). This shows that, up to a small multiplicative factor, the desired error bound (47) holds uniformly in each sub-interval with overwhelming probability. Finally, thanks to the overwhelming probability, a union bound over the n^2 sub-intervals yields the desired uniform estimate (47) for all $z = E + \mathrm{i}\eta$ with $|E - \lambda_{\mathrm{R}}| \le n^{-2/3+\delta}$ and $\eta = n^{-2/3-\delta}$.

Using the same arguments, we can prove a similar bound for the $\mathbf{R}^{[k]}_{\alpha\beta}$ and $\mathbf{R}_{\alpha\beta}$ blocks. Hence, we have shown the desired results. □

D.3 Stability of the top eigenvalue

As a consequence of the stability of the resolvent, we also have the stability of the top eigenvalue. This stability of the eigenvalue will play a crucial rule for the resolvent approximation of eigenvector statistics in the next subsection.

Lemma D.5. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. Assume $k \le n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$. Let $0 < \delta < \delta_0$ with δ_0 as in Lemma D.4. For any $\varepsilon > 0$, with overwhelming probability, we have*

$$\left|\lambda - \lambda^{[k]}\right| \le n^{-\frac{2}{3}-\delta+\varepsilon}.$$

Proof. Without loss of generality, we assume that $\lambda > \lambda^{[k]}$. Set $\eta = n^{-2/3-\delta}$. By the spectral representation of the resolvent (41), we have

$$\mathrm{Im}\,\mathbf{R}_{ii}(z) = \eta \sum_{\ell=1}^{p} \frac{|\mathbf{v}_\ell(i)|^2}{(\lambda_\ell - E)^2 + \eta^2} \ge \frac{\eta|\mathbf{v}(i)|^2}{(\lambda - E)^2 + \eta^2} \ge \frac{\eta|\mathbf{v}(i)|^2}{2\left(\max(|\lambda - E|, \eta)\right)^2}.$$

By the pigeonhole principle, we know that there exists $1 \le i \le p$ such that $|\mathbf{v}(i)| \ge p^{-1/2}$. Choosing this i and $z = \lambda + \mathrm{i}\eta$, we obtain

$$p\eta^{-1}\mathrm{Im}\,\mathbf{R}_{ii}(\lambda + \mathrm{i}\eta) \ge \frac{1}{2\eta^2}. \tag{61}$$

On the other hand, using the spectral representation of resolvent again for $\mathbf{R}^{[k]}$, we have

$$p\eta^{-1}\mathrm{Im}\,\mathbf{R}^{[k]}_{jj}(z) = \sum_{m=1}^{p} \frac{p|\mathbf{v}^{[k]}_m(j)|^2}{\left(\lambda^{[k]}_m - \lambda\right)^2 + \eta^2}.$$

Pick $\omega > 0$, we decompose the summation into two parts

$$J_1 = \sum_{m=1}^{n^\omega} \frac{p|\mathbf{v}^{[k]}_m(j)|^2}{\left(\lambda^{[k]}_m - \lambda\right)^2 + \eta^2}, \quad J_2 = \sum_{m=n^\omega+1}^{p} \frac{p|\mathbf{v}^{[k]}_m(j)|^2}{\left(\lambda^{[k]}_m - \lambda\right)^2 + \eta^2}.$$

Using delocalization of eigenvectors, for any $\varepsilon > 0$, with overwhelming probability, we have

$$J_1 \lesssim \frac{n^{\omega+\varepsilon}}{(\min_{1\le m\le p} |\lambda_m^{[k]} - \lambda|)^2}. \tag{62}$$

By the Tracy-Widom limit of the top eigenvalue (Lemma B.2), for any $\varepsilon > 0$, with overwhelming probability, we have $|\lambda - \lambda_{\mathrm{R}}| \le n^{-2/3+\varepsilon}$. Also, as discussed in (19), the rigidity of eigenvalues yields that for all $m \ge n^{\omega}$, with overwhelming probability,

$$\lambda - \lambda_m^{[k]} \gtrsim m^{2/3} p^{-2/3}.$$

Then using delocalization again, with overwhelming probability, we have

$$J_2 \le \sum_{m=n^{\omega}+1}^{p} \frac{n^{\varepsilon}}{(\lambda_m^{[k]} - \lambda)^2} \lesssim n^{\varepsilon} (n^{\omega})^{-1/3} n^{4/3}. \tag{63}$$

Again, since $|\lambda^{[k]} - \lambda| \le 2n^{-2/3+\varepsilon}$, by choosing $\omega = 2\varepsilon$ we have $J_2 \le J_1$. Therefore, by (62) and (63), we have shown that with overwhelming probability

$$p\eta^{-1} \mathrm{Im}\, \mathbf{R}_{jj}^{[k]}(\lambda + \mathrm{i}\eta) \lesssim n^{3\varepsilon} \left(\min_{1\le m\le p} |\lambda_m^{[k]} - \lambda| \right)^{-2}.$$

Note that the minimum is attained by $\lambda^{[k]}$. This shows that

$$n\eta^{-1} \mathrm{Im}\, \mathbf{R}_{ii}^{[k]}(\lambda + \mathrm{i}\eta) \lesssim n^{3\varepsilon} |\lambda^{[k]} - \lambda|^{-2}.$$

Using Lemma D.4 and (61), we have

$$n\eta^{-1} \mathrm{Im}\, \mathbf{R}_{ii}^{[k]}(\lambda+\mathrm{i}\eta) \ge n\eta^{-1} \left(\mathrm{Im}\, \mathbf{R}_{ii}(\lambda + \mathrm{i}\eta) - \left| \mathrm{Im}\, \mathbf{R}_{ii}^{[k]}(\lambda + \mathrm{i}\eta) - \mathrm{Im}\, \mathbf{R}_{ii}(\lambda + \mathrm{i}\eta) \right| \right) \ge \frac{1}{2\eta^2} - \frac{1}{n^c \eta^2} \gtrsim \frac{1}{\eta^2}.$$

Therefore, we have shown that, with overwhelming probability,

$$\frac{1}{\eta^2} \lesssim n^{3\varepsilon} \frac{1}{|\lambda - \lambda^{[k]}|^2}.$$

Recall $\eta = n^{-2/3-\delta}$, and we conclude that

$$\left| \lambda - \lambda^{[k]} \right| \le n^{-2/3-\delta+3\varepsilon},$$

which proves the desired result thanks to the arbitrariness of $\varepsilon > 0$. □

D.4 Proof of Theorem 1.2

The final ingredient to prove the resampling stability is the following approximation lemma, which asserts that the product of entries in the eigenvector can be well approximated by the resolvent entries.

Lemma D.6. *Consider the weakly spiked model $\mathcal{O} = \emptyset$. Assume that $k \ll n^{5/3-\epsilon_0}$ for some $\epsilon_0 > 0$. Let $0 < \delta < \delta_0$ be as in Lemma D.4. Then, for $z_0 = \lambda + \mathrm{i}\eta$ with $\eta = n^{-2/3-\delta}$, there exists $c' > 0$ such that with probability $1 - o(1)$ we have*

$$\max_{i,j} |\eta \mathrm{Im}\, \mathbf{R}_{ij}(z_0) - \mathbf{v}(i)\mathbf{v}(j)| \le n^{-1-c'}, \quad \textit{and} \quad \max_{i,j} \left| \eta \mathrm{Im}\, \mathbf{R}_{ij}^{[k]}(z_0) - \mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j) \right| \le n^{-1-c'}.$$

Similarly, we also have

$$\max_{\alpha,\beta}\left|\eta\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}-\mathbf{u}(\alpha)\mathbf{u}(\beta)\right|\le n^{-1-c'},\quad \textit{and}\quad \max_{\alpha,\beta}\left|\eta\mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right|\le n^{-1-c'}.$$

Proof. For any $\varepsilon>0$, we consider a general $z=E+\mathrm{i}\eta$ with $|E-\lambda_{\mathrm{R}}|\le n^{-2/3+\varepsilon}$. From the spectral representation of the resolvent (41), we have

$$\mathrm{Im}\,\mathbf{R}_{ij}(z)=\eta\sum_{\ell=1}^{p}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2}.$$

Pick some $\omega>0$, we decompose the summation on the right-hand side into three parts

$$\sum_{\ell=1}^{p}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2}=\frac{\mathbf{v}(i)\mathbf{v}(j)}{(\lambda-E)^2+\eta^2}+J_1+J_2,$$

where

$$J_1=\sum_{\ell=2}^{n^\omega}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2},\quad J_2=\sum_{\ell=n^\omega+1}^{p}\frac{\mathbf{v}_\ell(i)\mathbf{v}_\ell(j)}{(\lambda_\ell-E)^2+\eta^2}.$$

Using the same arguments as in (63), for any $\varepsilon>0$, with overwhelming probability we have

$$|J_2|\lesssim n^{\varepsilon}(n^{\omega})^{-1/3}n^{1/3}.$$

For the term J_1, we consider the following event

$$\mathcal{E}:=\left\{\lambda_1-\lambda_2\ge c_0n^{-2/3}\right\}\cap\left\{\max_{1\le\ell\le p}\|\mathbf{v}_\ell\|_\infty\le n^{-1/2+\varepsilon}\right\}\cap\left\{|J_2|\lesssim n^{\varepsilon}(n^{\omega})^{-1/3}n^{4/3}\right\}.$$

For any $\varepsilon>0$, we can find an appropriate $c_0>0$ such that $\mathbb{P}(\mathcal{E})>1-\varepsilon/2$. Then, for $z=E+\mathrm{i}\eta$ with $|\lambda-E|\le\frac{c_0}{2}n^{-2/3}$, on the event $\mathcal{E}$, we have

$$|J_1|\lesssim n^{\varepsilon}n^{\omega}n^{1/3}.$$

Let $\delta'>0$ with $\delta'+\delta<\delta_0$. On the event $\mathcal{E}$, for all $z=E+\mathrm{i}\eta$ with $|\lambda-E|\le\eta n^{-\delta'}$ and $\eta=n^{-2/3-\delta}$, we have

$$\left|\mathbf{v}(i)\mathbf{v}(j)-\frac{\eta^2\mathbf{v}(i)\mathbf{v}(j)}{(\lambda-E)^2+\eta^2}\right|\le n^{-1+2\varepsilon}\left|1-\frac{\eta^2}{(\lambda-E)^2+\eta^2}\right|\le n^{-1+2\varepsilon-2\delta'}.$$

This yields

$$\begin{aligned}|\eta\mathrm{Im}\,\mathbf{R}_{ij}(z)-\mathbf{v}(i)\mathbf{v}(j)|&\le\left|\mathbf{v}(i)\mathbf{v}(j)-\frac{\eta^2\mathbf{v}(i)\mathbf{v}(j)}{(\lambda-E)^2+\eta^2}\right|+\eta^2(|J_1|+|J_2|)\\&\le n^{-1+2\varepsilon-2\delta'}+n^{-1+\varepsilon+\omega-2\delta}+n^{-1+\varepsilon-\frac{\omega}{3}-2\delta}.\end{aligned}$$

Choosing $\omega=\varepsilon<\min(\delta,\delta')/2$, we obtain

$$\max_{i,j}|\eta\mathrm{Im}\,\mathbf{R}_{ij}(z)-\mathbf{v}(i)\mathbf{v}(j)|\le n^{-1-\min(\delta,\delta')}.\tag{64}$$

Similarly, we can apply the same arguments to $\mathbf{R}^{[k]}$. Consider the event

$$\mathcal{E}':=\left\{\max_{i,j}\left|\eta\mathrm{Im}\,\mathbf{R}^{[k]}_{ij}(z)-\mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right|\le n^{-1-\min(\delta,\delta')}\text{ for all }|\lambda^{[k]}-E|\le\eta n^{-\delta'},\eta=n^{-2/3-\delta}\right\}.$$

By previous arguments, we know $\mathbb{P}(\mathcal{E}') > 1-\varepsilon/2$. This gives us $\mathbb{P}(\mathcal{E}\cap\mathcal{E}') > 1-\varepsilon$. Finally, note that $\delta+\delta' < \delta_0$, by Lemma D.5, with overwhelming probability we have $|\lambda-\lambda^{[k]}| \le n^{-2/3-\delta-\delta'} = \eta n^{-\delta'}$. This implies that we can choose $z=\lambda+\mathrm{i}\eta$ in both (64) and $\mathcal{E}'$. Thus, we have shown the desired result for $\mathbf{v}$ and $\mathbf{v}^{[k]}$ by choosing $0 < c' < \min(\delta,\delta')$.

On the other hand, from (41) we also have

$$\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z)}{z} = \eta\sum_{\ell=1}^{n}\frac{\mathbf{u}_\ell(\alpha)\mathbf{u}_\ell(\beta)}{(\lambda_\ell-E)^2+\eta^2}.$$

Using the same methods as above yields the desired result for $\mathbf{u}$ and $\mathbf{u}^{[k]}$. □

Now we prove the main result Theorem 1.2 on the stability of PCA under moderate resampling for the weakly spiked model.

Proof of Theorem 1.2. Let $z_0=\lambda+\mathrm{i}\eta$ as in Lemma D.6. By Lemma D.4 and D.6, we know that, with probability $1-o(1)$, for all $\alpha,\beta\in\mathcal{I}_2$, we have

$$\begin{aligned}
&\left|\mathbf{v}(i)\mathbf{v}(j)-\mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right|\\
&\le |\mathbf{v}(i)\mathbf{v}(j)-\eta\mathrm{Im}\,\mathbf{R}_{ij}(z_0)| + \left|\eta\mathrm{Im}\,\mathbf{R}_{ij}(z_0)-\eta\mathrm{Im}\,\mathbf{R}^{[k]}_{ij}(z_0)\right| + \left|\eta\mathrm{Im}\,\mathbf{R}^{[k]}_{ij}(z_0)-\mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right|\\
&\le n^{-1-c}+n^{-1-c'}+n^{-1-c}.
\end{aligned}$$

Denote $c'' := \min(c,c')$, and we have

$$\max_{i,j}\left|\mathbf{v}(i)\mathbf{v}(j)-\mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \lesssim n^{-1-c''}.$$

For any fixed $\varepsilon>0$, we consider the event

$$\mathcal{E} := \left\{\max_{i,j}\left|\mathbf{v}(i)\mathbf{v}(j)-\mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right| \lesssim n^{-1-c''}\right\}\cap\left\{\|\mathbf{v}^{[k]}\|_\infty \le n^{-1/2+\varepsilon}\right\}.$$

Since delocalization of eigenvectors holds with overwhelming probability, we know that $\mathbb{P}(\mathcal{E})=1-o(1)$.

By the pigeonhole principle, there exists $1\le i\le p$ such that $|\mathbf{v}(i)| > p^{-1/2}$. We choose the $\pm$ phases of $\mathbf{v}$ and $\mathbf{v}^{[k]}$ in the way that $\mathbf{v}(i)$ and $\mathbf{v}^{[k]}(i)$ are non-negative. On the event $\mathcal{E}$, we obtain

$$\left|\mathbf{v}(i)-\mathbf{v}^{[k]}(i)\right| = \frac{\left|(\mathbf{v}(i))^2-(\mathbf{v}^{[k]}(i))^2\right|}{\mathbf{v}(i)+\mathbf{v}^{[k]}(i)} \lesssim n^{-1/2-c''}.$$

Moreover, for any entry $\mathbf{v}(j)$ and $\mathbf{v}^{[k]}(j)$, if $\mathcal{E}$ holds, the triangle inequality gives us

$$\begin{aligned}
\left|\mathbf{v}(j)-\mathbf{v}^{[k]}(j)\right| &= \frac{\left|\mathbf{v}(i)\mathbf{v}(j)-\mathbf{v}(i)\mathbf{v}^{[k]}(j)\right|}{\mathbf{v}(i)}\\
&\le \frac{\left|\mathbf{v}(i)\mathbf{v}(j)-\mathbf{v}^{[k]}(i)\mathbf{v}^{[k]}(j)\right|}{\mathbf{v}(i)} + \frac{|\mathbf{v}^{[k]}(j)|}{\mathbf{v}(i)}|\mathbf{v}(i)-\mathbf{v}^{[k]}(i)|\\
&\lesssim n^{-1/2-c''}+n^{-1/2-c''+\varepsilon}.
\end{aligned}$$

Choosing ε sufficiently small, this implies the desired result.

For $\mathbf{u}$ and $\mathbf{u}^{[k]}$, note that

$$\begin{aligned}&\left|\mathbf{u}(\alpha)\mathbf{u}(\beta)-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right|\\ \le&\left|\mathbf{u}(\alpha)\mathbf{u}(\beta)-\eta\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}\right|+\left|\eta\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}-\eta\mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}\right|+\left|\eta\mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right|\end{aligned}$$

By Lemma D.4, we have

$$\left|\mathrm{Im}\,\frac{\mathbf{R}_{\alpha\beta}(z_0)}{z_0}-\mathrm{Im}\,\frac{\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}\right|\le\left|\frac{\mathbf{R}_{\alpha\beta}(z_0)-\mathbf{R}^{[k]}_{\alpha\beta}(z_0)}{z_0}\right|\lesssim\left|\mathbf{R}_{\alpha\beta}(z_0)-\mathbf{R}^{[k]}_{\alpha\beta}(z_0)\right|\le\frac{1}{n^{1+c}\eta}.$$

As a consequence, we have

$$\left|\mathbf{u}(\alpha)\mathbf{u}(\beta)-\mathbf{u}^{[k]}(\alpha)\mathbf{u}^{[k]}(\beta)\right|\lesssim n^{-1-c''}.$$

The desired result for $\mathbf{u}$ and $\mathbf{u}^{[k]}$ then follows from the same arguments above for $\mathbf{v}$ and $\mathbf{v}^{[k]}$. □

E Proof for the Strongly Spiked Model

As discussed in Section B, one of the key differences between the weakly spiked model and the strongly spiked model is the distribution of eigenvectors. We see from the previous sections that the proof for the weakly spiked model crucially depends on the delocalization property. In contrast, this is not valid in the strongly spiked case and consequently results in a distinct phenomenon.

In the strongly spiked model, the celebrated BBP phase transition [BBAP05] shows that the leading sample eigenvectors in the outlier of the spectrum have non-trivial correlation with the corresponding population eigenvectors. Recall that the population covariance matrix is in the form $\mathbf{\Sigma}=\sum_{i=1}^p d_i\mathbf{e}_i\mathbf{e}_i^\top$, and indices i with $d_i>1+\sqrt{\xi}$ correspond to the outlier (denoted as $\mathcal{O}$). For $i\in\mathcal{O}$, it was first derived in [Lu02] and later generalized in [JL09, BGN11] that

$$|\langle\mathbf{v}_i,\mathbf{e}_i\rangle|^2=\frac{1-\frac{\xi}{(d_i-1)^2}}{1+\frac{\xi}{d_i-1}}+o(1),\quad a.s. \tag{65}$$

Since $\mathbf{X}$ and $\mathbf{X}^{[k]}$ have the same marginal distribution, the same also holds for $\mathbf{v}_i^{[k]}$. Note that the eigenvector overlap $|\langle\mathbf{v},\mathbf{v}^{[k]}\rangle|$ is independent of the sign the of principal components. Therefore, without loss of generality, we may assume that

$$\langle\mathbf{v}_i,\mathbf{e}_i\rangle=\langle\mathbf{v}_i^{[k]},\mathbf{e}_i\rangle=\sqrt{\frac{1-\frac{\xi}{(d_i-1)^2}}{1+\frac{\xi}{d_i-1}}}+o(1)$$

Since both principal components $\mathbf{v}$ and $\mathbf{v}^{[k]}$ lie on the unit sphere, we obtain

$$\|\mathbf{v}_i-\mathbf{e}_i\|^2=2-2\langle\mathbf{v}_i,\mathbf{e}_i\rangle=2-2\sqrt{\frac{1-\frac{\xi}{(d_i-1)^2}}{1+\frac{\xi}{d_i-1}}}+o(1)$$

and same also holds for $\mathbf{v}_i^{[k]}$. By triangle inequality,

$$\|\mathbf{v}-\mathbf{v}^{[k]}\| \le \|\mathbf{v}_i-\mathbf{e}_i\| + \|\mathbf{v}_i^{[k]}-\mathbf{e}_i\| \le 2\sqrt{2-2\sqrt{\frac{1-\frac{\xi}{(d_i-1)^2}}{1+\frac{\xi}{d_i-1}}}} + o(1)$$

Hence,

$$|\langle \mathbf{v}_i, \mathbf{v}_i^{[k]}\rangle| = \frac{2-\|\mathbf{v}_i-\mathbf{v}_i^{[k]}\|^2}{2} \ge 1-4\left(1-\sqrt{\frac{1-\frac{\xi}{(d_i-1)^2}}{1+\frac{\xi}{d_i-1}}}\right) + o(1)$$

which completes the proof.

F More Discussions on Database Alignment

Recall that in the database alignment problem, we have two matrices $\mathbf{X} \in \mathbb{R}^{p\times n}, \quad \mathbf{Y} = \mathbf{\Pi}_{\mathrm{f}}\mathbf{X}^{[k]}\mathbf{\Pi}_{\mathrm{s}}^\top$ where $\mathbf{\Pi}_{\mathrm{s}}$ and $\mathbf{\Pi}_{\mathrm{f}}$ are permutation matrices of order n and p chosen uniformly at random. The goal is to recover the permutations $\mathbf{\Pi}_{\mathrm{s}}$ and $\mathbf{\Pi}_{\mathrm{f}}$ based on the observations $\mathbf{X}$ and $\mathbf{Y}$. To separate the sample permutation $\mathbf{\Pi}_{\mathrm{s}}$ and the feature permutation $\mathbf{\Pi}_{\mathrm{f}}$, we consider

$$\mathbf{A} = \mathbf{X}\mathbf{X}^\top, \quad \mathbf{B} = \mathbf{Y}\mathbf{Y}^\top = \mathbf{\Pi}_{\mathrm{f}}\left(\mathbf{X}^{[k]}(\mathbf{X}^{[k]})^\top\right)\mathbf{\Pi}_{\mathrm{f}}^\top,$$

and

$$\widehat{\mathbf{A}} = \mathbf{X}^\top\mathbf{X}, \quad \widehat{\mathbf{B}} = \mathbf{Y}^\top\mathbf{Y} = \mathbf{\Pi}_{\mathrm{s}}\left((\mathbf{X}^{[k]})^\top\mathbf{X}^{[k]}\right)\mathbf{\Pi}_{\mathrm{s}}^\top.$$

In database alignment or graph matching, spectral methods have been studied and applied in many scenarios (see e.g. [FMWX20, FMWX22] and [GLM22]), and one of the most common spectral algorithm is to focus on the top eigenvectors. In this manner, a natural idea to reconstruct the permutations $\mathbf{\Pi}_{\mathrm{s}}$ (and $\mathbf{\Pi}_{\mathrm{f}}$) in our setup is to align the top eigenvectors of the matrices $\mathbf{A}$ and $\mathbf{B}$ (and $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$). See Algorithm 1 for details, and note that apparently this algorithm is computationally efficient. We are interested in under what resampling strength, the `PCA-Recovery` algorithm can almost perfectly reconstruct the permutations, and under what condition this method completely fail.

A similar PCA method was studied in [GLM22] to match two symmetric Gaussian matrices correlated via additive Gaussian noise. Their work proved a all-or-nothing phenomenon in the alignment problem (i.e. the accuracy of the recovery undergoes a sharp transition from 0 to 1 near some critical threshold), and a key step of their proof is a 0-1 transition for the inner product of the top eigenvectors. Since our weakly spiked model has the same phase transition for the eigenvector overlap, it is natural to ask if our alignment problem for the weakly spiked model also exhibits the all-or-nothing phenomenon. However, the arguments in [GLM22] are not applicable in our case. Their proof heavily depends on the Gaussian assumption of the matrices, and the additive strucutre of the noise. In particular, they proof crucially relies on the orthogonal invariance of the Gaussian noise. While in our case, the noise is presented in terms of the resampling strength. There is no way to write the “noise" in an additive form that is independent of the “signal". Even in the case of Gaussian null model, a rigorous analysis of the `PCA-Recovery` algorithm seems difficult.

Nevertheless, our results on the sensitivity of the eigenvector inner products suggest that, in the weakly spiked model, when $k \gg n^{5/3}$, the two eigenvectors are approximately de-correlated so that they share almost no common information. Consequently, recovery of the data via aligning the principal components would be basically random guessing. Therefore, in the weakly spiked case,

Algorithm 1 PCA-Recovery

Input: data matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times p}$
Output: permutation matrices $\widehat{\mathbf{\Pi}}_{\mathrm{s}} \in \mathbb{R}^{n \times n}$, $\widehat{\mathbf{\Pi}}_{\mathrm{f}} \in \mathbb{R}^{p \times p}$
Compute $\mathbf{u}$ the unit leading left singular vectors of $\mathbf{X}$
Compute $\mathbf{v}$ the unit leading right singular vectors of $\mathbf{X}$
Compute $\mathbf{u}'$ the unit leading left singular vectors of $\mathbf{Y}$
Compute $\mathbf{v}'$ the unit leading right singular vectors of $\mathbf{Y}$

Compute $\mathbf{\Pi}_{\mathrm{s}}^{+}$ the permutation aligning $\mathbf{u}$ and $\mathbf{u}'$
Compute $\mathbf{\Pi}_{\mathrm{s}}^{-}$ the permutation aligning $\mathbf{u}$ and $-\mathbf{u}'$
Compute $\mathbf{\Pi}_{\mathrm{f}}^{+}$ the permutation aligning $\mathbf{v}$ and $\mathbf{v}'$
Compute $\mathbf{\Pi}_{\mathrm{f}}^{-}$ the permutation aligning $\mathbf{v}$ and $-\mathbf{v}'$

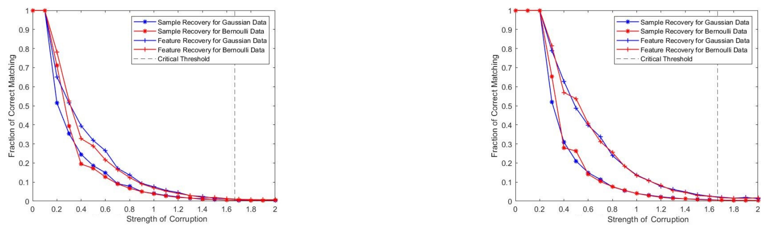
if $\langle \mathbf{A}, \mathbf{\Pi}_{\mathrm{s}}^{+}\mathbf{B}(\mathbf{\Pi}_{\mathrm{s}}^{+})^{\top} \rangle \geq \langle \mathbf{A}, \mathbf{\Pi}_{\mathrm{s}}^{-}\mathbf{B}(\mathbf{\Pi}_{\mathrm{s}}^{-})^{\top} \rangle$ **then**
  $\widehat{\mathbf{\Pi}}_{\mathrm{s}} \leftarrow \mathbf{\Pi}_{\mathrm{s}}^{+}$
else
  $\widehat{\mathbf{\Pi}}_{\mathrm{s}} \leftarrow \mathbf{\Pi}_{\mathrm{s}}^{-}$
end if

if $\langle \widehat{\mathbf{A}}, \mathbf{\Pi}_{\mathrm{f}}^{+}\widehat{\mathbf{B}}(\mathbf{\Pi}_{\mathrm{f}}^{+})^{\top} \rangle \geq \langle \widehat{\mathbf{A}}, \mathbf{\Pi}_{\mathrm{f}}^{-}\widehat{\mathbf{B}}(\mathbf{\Pi}_{\mathrm{f}}^{-})^{\top} \rangle$ **then**
  $\widehat{\mathbf{\Pi}}_{\mathrm{f}} \leftarrow \mathbf{\Pi}_{\mathrm{f}}^{+}$
else
  $\widehat{\mathbf{\Pi}}_{\mathrm{f}} \leftarrow \mathbf{\Pi}_{\mathrm{f}}^{-}$
end if

we conjecture that if $k \gg n^{5/3}$, `PCA-Recovery` fails to recover the latent permutations in the sense that it can only achieve $o(1)$ fraction of correct matching with the ground truth. On the other hand, when $k \ll n^{5/3}$, the performance of our algorithm seems mysterious.

As shown in Figure 3, we empirically check the performance of the `PCA-Recovery` algorithm for the null model and the weakly spiked model. Similarly as in setup in the Main Part, we consider a data matrix of size 250×1000 whose entries are Gaussian or two-point, and the population covariance matrix is either identity or contains weak spikes of rank $r = 10$ with strength $\{d_i\}_{i=1}^{10}$ uniformly sampled from $(1, \frac{3}{2})$. Numerical simulations suggest that when $k \gg n^{5/3}$, the performance of `PCA-Recovery` is indeed poor in the sense that the accuracy of the recovery is almost 0. On the other hand, when $k \ll n^{5/3}$, experiments show that we cannot expect the sharp all-or-nothing phenomenon similarly as in [GLM22].

Finally, we remark that what `PCA-Recovery` actually studies is a more difficult task, as we do not need direct observations of $\mathbf{X}$ and $\mathbf{Y}$. We can consider a harder problem (in both statistical and computational sense), which we call alignment from covariance profile. In this problem, we only have access to the covariance between the samples and we aim to recover the correspondence between the samples from the two databases. A similar problem with Gaussian data and additive noise was considered in [WWXY22] as a prototype for matching random geometric dot-product graphs. The analysis of such database alignment problem with adversarial corruption will be an interesting direction for future studies.



(a) Null model

(b) Weakly spiked model

Figure 3: Recovery accuracy for 250×1000 matrices with Gaussian and two-point data. The horizontal axis is the corruption (resampling) strength, given by $\log_n(4k)$. Each experiment is averaged over 50 repetitions.

References

[AEK14] Oskari Ajanki, Lászlo Erdős, and Torben Krüger. Local semicircle law with imprimitive variance matrix. *Electronic Communications in Probability*, 19:1–9, 2014.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge university press, 2010.

[ASX17] Yacine Ait-Sahalia and Dacheng Xiu. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399, 2017.

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.

[BCRT21] Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507, 2021.

[BDMN05] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138, 2005.

[BE02] Olivier Bousquet and André Elisseeff. Stability and generalization. *The Journal of Machine Learning Research*, 2:499–526, 2002.

[BEK+14] Alex Bloemendal, László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19:no. 33, 53, 2014.

[BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

[BKS99] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, (90):5–43 (2001), 1999.

[BKYY16] Alex Bloemendal, Antti Knowles, Horng-Tzer Yau, and Jun Yin. On the principal components of sample covariance matrices. *Probab. Theory Related Fields*, 164(1-2):459–552, 2016.

[BL22] Charles Bordenave and Jaehun Lee. Noise sensitivity for the top eigenvector of a sparse random matrix. *Electronic Journal of Probability*, 27:1–50, 2022.

[BLZ20] Charles Bordenave, Gábor Lugosi, and Nikita Zhivotovskiy. Noise sensitivity of the top eigenvector of a Wigner matrix. *Probability Theory and Related Fields*, 177(3):1103–1135, 2020.

[BS06] Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408, 2006.

[CBG16] Romain Couillet and Florent Benaych-Georges. Kernel spectral clustering of large dimensional data. *Electronic Journal of Statistics*, 10(1):1393–1454, 2016.

[Cha05] Sourav Chatterjee. Concentration inequalities with exchangeable pairs. page 105, 2005. Thesis (Ph.D.)–Stanford University.

[Cha07] Sourav Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007.

[CLMW11] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

[CS13] Xiuyuan Cheng and Amit Singer. The spectrum of random inner-product kernel matrices. *Random Matrices: Theory and Applications*, 2(04):1350010, 2013.

[CSS13] Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. A near-optimal algorithm for differentially-private principal components. *Journal of Machine Learning Research*, 14, 2013.

[DCK19] Osman E Dai, Daniel Cullina, and Negar Kiyavash. Database alignment with Gaussian features. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3225–3233. PMLR, 2019.

[DHS21] Zhun Deng, Hangfeng He, and Weijie Su. Toward better generalization bounds with locally elastic stability. In *International Conference on Machine Learning*, pages 2590–2600. PMLR, 2021.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[DT11] Momar Dieng and Craig A Tracy. Application of random matrix theory to multivariate statistics. In *Random Matrices, Random Processes and Integrable Systems*, pages 443–507. Springer, 2011.

[DY18] Xiucai Ding and Fan Yang. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *The Annals of Applied Probability*, 28(3):1679–1738, 2018.

[EEPK05] Andre Elisseeff, Theodoros Evgeniou, Massimiliano Pontil, and Leslie Pack Kaelbing. Stability of randomized learning algorithms. *Journal of Machine Learning Research*, 6(1), 2005.

[EK10] Noureddine El Karoui. The spectrum of kernel random matrices. *Ann. Statist.*, 38(1):1–50, 2010.

[FMWX20] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In *International conference on machine learning*, pages 2985–2995. PMLR, 2020.

[FMWX22] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II. *Foundations of Computational Mathematics*, pages 1–51, 2022.

[FWZ18] Jianqing Fan, Weichen Wang, and Yiqiao Zhong. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18(207):1–42, 2018.

[GLM22] Luca Ganassali, Marc Lelarge, and Laurent Massoulié. Spectral alignment of correlated gaussian matrices. *Advances in Applied Probability*, 54(1):279–310, 2022.

[GS14] Christophe Garban and Jeffrey E Steif. *Noise sensitivity of Boolean functions and percolation*, volume 5. Cambridge University Press, 2014.

[HPY18] Xiao Han, Guangming Pan, and Qing Yang. A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli*, 24(4B):3447–3468, 2018.

[HPZ16] Xiao Han, Guangming Pan, and Bo Zhang. The Tracy–Widom law for the largest eigenvalue of F type matrices. *The Annals of Statistics*, 44(4):1564–1592, 2016.

[HRS16] Moritz Hardt, Ben Recht, and Yoram Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *International conference on machine learning*, pages 1225–1234. PMLR, 2016.

[JCL21] Ji Hyung Jung, Hye Won Chung, and Ji Oon Lee. Detection of signal in the spiked rectangular models. In *International Conference on Machine Learning*, pages 5158–5167. PMLR, 2021.

[JL09] Iain M Johnstone and Arthur Yu Lu. Sparse principal components analysis. *arXiv preprint arXiv:0901.4392*, 2009.

[JN17] Iain M Johnstone and Boaz Nadler. Roy's largest root test under rank-one alternatives. *Biometrika*, 104(1):181–193, 2017.

[Joh01] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327, 2001.

[Joh07] Iain M. Johnstone. High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I*, pages 307–333. Eur. Math. Soc., Zürich, 2007.

[KB21] Byol Kim and Rina Foygel Barber. Black box tests for algorithmic stability. *arXiv preprint arXiv:2111.15546*, 2021.

[KN02] Samuel Kutin and Partha Niyogi. Almost-everywhere algorithmic stability and generalization error. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 275–282, 2002.

[LR10] Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electronic Journal of Probability*, 15:1319–1343, 2010.

[LS16] Ji Oon Lee and Kevin Schnelli. Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.*, 26(6):3786–3839, 2016.

[Lu02] Arthur Yu Lu. *Sparse principal component analysis for functional data.* Stanford University, 2002.

[MNPR06] Sayan Mukherjee, Partha Niyogi, Tomaso Poggio, and Ryan Rifkin. Learning theory: stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization. *Advances in Computational Mathematics*, 25(1):161–193, 2006.

[NJW01] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.

[Pau07] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[PWBM18] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca i: Spiked random matrix models. *The Annals of Statistics*, 46(5):2416–2451, 2018.

[PY14] Natesh S. Pillai and Jun Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001, 2014.

[Rin08] Markus Ringnér. What is principal component analysis? *Nature biotechnology*, 26(3):303–304, 2008.

[SX21] Kevin Schnelli and Yuanyuan Xu. Convergence rate to the Tracy–Widom laws for the largest eigenvalue of sample covariance matrices. *arXiv preprint arXiv:2108.02728*, 2021.

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

[TV12] Terence Tao and Van Vu. Random covariance matrices: universality of local statistics of eigenvalues. *Ann. Probab.*, 40(3):1285–1315, 2012.

[VK06] Seema Vyas and Lilani Kumaranayake. Constructing socio-economic status indices: how to use principal components analysis. *Health policy and planning*, 21(6):459–468, 2006.

[VL07] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17:395–416, 2007.

[Wan12] Ke Wang. Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices: Theory and Applications*, 1(01):1150005, 2012.

[Wan19] Haoyu Wang. Quantitative universality for the largest eigenvalue of sample covariance matrices. *arXiv preprint arXiv:1912.05473*, 2019.

[Wan22] Haoyu Wang. Optimal smoothed analysis and quantitative universality for the smallest singular value of random matrices. *arXiv preprint arXiv:2211.03975*, 2022.

[WWXY22] Haoyu Wang, Yihong Wu, Jiaming Xu, and Israel Yolou. Random graph matching in geometric models: the case of complete graphs. In *Conference on Learning Theory*, pages 3441–3488. PMLR, 2022.

[WXS22] Yihong Wu, Jiaming Xu, and H Yu Sophie. Settling the sharp reconstruction thresholds of random graph matching. *IEEE Transactions on Information Theory*, 2022.

[XCS10] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust pca via outlier pursuit. *Advances in neural information processing systems*, 23, 2010.

[Yan22] Fan Yang. Sample canonical correlation coefficients of high-dimensional random vectors: Local law and Tracy–Widom limit. *Random Matrices: Theory and Applications*, 11(01):2250007, 2022.